Spatial Analysis of Teen Driver-Related Crashes in Kansas

By

Hemin Jalal Mohammed
Ph.D., University Of Kansas
M.Sc., University Of Kansas, 2015
B.Sc., University of Tikrit, 2000

Submitted to the graduate degree program in Civil, Environmental and Architectural Engineering and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chair: Steven D. Schrock

Jie Han

Bonnie J. Johnson

Alexandra Kondyli

Thomas E. Mulinazzi

Date Defended: 29 May 2019

The dissertation committee for Hemin Jalal Mohammed certifies that this is the approved version of the following dissertation:

Spatial Analysis of Teen Driver-Related Crashes in Kansas

Chair: Steven D. Schrock

Date Approved: 29 May 2019

ABSTRACT

In the United States, the risk of vehicle crashes is higher among teens than among any other age group. Accordingly, in Kansas, teen drivers ages 14 to 19 were one of the primary foci of Kansas Department of Transportation's (KDOT's) Strategic Highway Safety Plan (SHSP) to reduce the number of traffic injuries and fatalities. However, after several years of improving metrics, it appears that overall teen crashes have begun to increase in the past few years and the SHSP goals were not met.

Most previous studies investigated the effects of demographic differences and nonspatial factors associated with crashes such as gender, age, driving under the influence of drugs or alcohol, the presence of passengers, and distractions. Besides these factors, it was necessary to investigate and understand how teen-related crashes are correlated and patterned spatially. However, adopting the spatial analysis methodology to identify the hotspots for teen drivers and factors behind their crashes has been underutilized.

This research was conducted to develop a methodology to identify statistically significant spatial patterns for crashes involving teen drivers. Also, modelling was performed that identified spatial relationships between teen-related crashes and contributed factors that significantly influence the number of these crashes using an ordinary linear regression (OLS) model and geographically weighted regression (GWR).

The utilized data were extracted from the KDOT crash database and other resources such as the Fatality Analysis Reporting System (FARS), the US Census Bureau, and the Kansas Department of Education. The analyzed crashes included crashes involving teen drivers aged between 14 and 19. The spatial analysis and modeling were conducted at the state level and Unified School District (USD) level using ArcGIS Pro software (Version 2.3.2).

The spatial analysis tools were used to find statistically significant hotspots and outliers for fatal and non-fatal crashes at the state level, and fatal and severe injuries at the USD level. Most of the statistically significant hotspots and outliers were centered in the most populated counties such as Johnson, Sedgwick, and Wyandotte County. From 18 candidate exploratory variables, two exploratory variables were statistically significant to build a predictive model using OLS and GWR. The two exploratory variables were the miles of non-state roads and the number of passenger cars in counties. The predictive model showed that the number of crashes involving teen driver was expected to be lower by more than three percent by 2026.

The methodology followed in this research was found to be applicable and valuable to spatially analyze teen-related crashes at the state and USD levels. The method was useful for analyzing a subset of crashes involving teen drivers; it can also be used to analyze other subgroups such as alcohol-related crashes, older driver crashes, or commercial vehicle crashes. The model represents useful guidance for the related parties' allocation of limited resource for reducing crashes, and is helpful in predicting future crashes based on historical trends.

DEDICATION

To my wife, Awaz.

To my children, Zhiro, Zhulyan, Vania, and Zeen.

ACKNOWLEDGMENTS

I owe my deepest gratitude to my supervisor professor Dr. Steven Schrock. He always guided me to improve the quality of my research. I could not have imagined having a better advisor for my graduate studies. He has been my best mentor, colleague, and friend along this thoughtful and rewarding journey.

I would like to express my gratitude to my dissertation committee for their support over the past years as they helped me to move from an idea to a completed study.

- Dr. Thomas E. Mulinazzi: Professor Emeritus in Civil, Environmental & Architectural Engineering Department, University of Kansas
- Dr. Jie Han: Professor in Civil, Environmental & Architectural Engineering Department, University of Kansas
- Dr. Bonnie J. Johnson: Associate Professor in Urban Planning Department, University of Kansas
- Dr. Alexandra Kondyli: Assistant Professor in Civil, Environmental & Architectural Engineering Department, University of Kansas

I would like to thank my life partner (Awaz) for her patience and her faith. You always supported my hustle and ambition all the way. Without you, this work would not have been completed.

I would also like to acknowledge my children (Zhiro, Zhulyan, Vania, and Zeen): You are the pride and joy of my life. I appreciate all your patience and support during my studies.

I would like to appreciate my parents and my siblings: You always believed in me and provided me with moral and emotional support in my life.

I further extend my gratitude to the Kansas Department of Transportation staff for providing access to the motor vehicle crash database and offering me the necessity data whenever I asked.

I would like to give a big thanks to Mr. Daniel H. Schulte for providing data on DVMT and number of vehicles registered in Kansas.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AADT – Annual Average Daily Traffic

AASHTO – American Association of State Highway Transportation Officials

ArcGIS – Aeronautical Reconnaissance Coverage Geographic Information System

ADT – Average Daily Traffic

CMF – Crash Modification Factors

CDC – Centers for Disease Control and Prevention

CSV –. Comma–Separated Values

CWOV – Collision with Other Vehicle

CRP – Continues Risk Profile

DMV – Department of Motor Vehicles

DUI – Driving Under the Influence

DVMT – Daily Vehicle Miles Traveled

EMS – Emergency Medical Services

EPDO – Equivalent Property Damage Only

ESP – Electronic Stability Program

ESRI – Environmental Systems Research Institute

FARS – Fatality Analysis Reporting System

FHWA – Federal Highway Administration

G_i* – Getis-Ord spatial statistics

GIS – Geographic Information System

GIS-T – Geographic Information Systems for Transportation

GDL – Graduated Driver Licensing

GWR – Geographically Weighted Regression

HP – Hazard Perception

HPT – Hazard Perception Test

HRL – Hazardous Road Locations

HSM – Highway Safety Manual

HSP – Highway Safety Plan

IIHS – Insurance Institute for Highway Safety

ISA – Intelligent Speed Adaptation

KDE – Kernel Density Estimation

KDOT – Kansas Department of Transportation

KSDE – Kansas Department of Education

NHTSA – National Highway Traffic Safety Administration

OECD – Organisation for Economic Co-operation and Development

OLS – Ordinary Least Squares

PDF – Portable Document Format

PDO – Property Damage Only

PS – Peak Searching

RTM – Regression to the Mean

SR – Simple Ranking

SW – Sliding Window

USD – Unified School District

USDOT – U.S. Department of Transportation

USGS – U.S. Geological Survey

VIF – Variance Inflation Factor

WHO – World Health Organization

CHAPTER I. INTRODUCTION

The Highway Safety Manual (HSM) refers to safety as "the crash frequency or crash severity, or both, and collision type for a specific time period, a given location, and a given set of geometric and operational conditions" (AASHTO, 2010). Furthermore, the crash is defined as "a set of events not under human control that results in injury or property damage due to the collision of at least one motorized vehicle and may involve a crash with another motorized vehicle, a bicyclist, a pedestrian or an object." In other words, a crash refers to anything that is hit by a vehicle (Evans, 2004). Crash analysis is a sufficient attempt to identify hazardous areas and causality to improve traffic safety.

Globally, around 1.25 million people died in 2013, as a result of road traffic crashes, and up to 50 million people sustained non-fatal injuries. Unfortunately, the number of road traffic deaths have increased by 13 percent from 2000 to 2013. Therefore, road traffic injuries represent a major threat for the world population, especially for teens, where it is the main cause of death among people aged 15–19 in the world (WHO, 2007; WHO, 2017). The root of this problem is not new. In the first annual international symposium of youth enhancement service in 1995, Simpson (1996) stated that the traffic crashes involving teen drivers aged 16 to 19 " have been a worldwide road safety and public health concern for several decades".

The United States is not excluded from this dilemma. In the United States, the risk of vehicle crashes is higher among teens than among any other age group (Lonero and Mayhew, 2010). The motor vehicle traffic crashes in the US are a leading cause of death for young people aged 16-20 years since 2001 (Hilton, 2006; Liu et al., 2015; Subramanian, 2012; Webb, 2016; NHTSA, 2018b; Subramanian, 2005). In 2015, 2,333 teens were killed meaning six teens died every day from road traffic injuries, and this number increased by 3.6 percent in 2016 (NHTSA, 2017). Per mile driven, teen drivers are nearly three times more likely to be in a fatal crash than drivers age 20 and older (IIHS, 2016). Hersman and Rosekind listed the factors that put young drivers at highest risk: driving under the influence (DUI) of alcohol and drugs, speeding, and seatbelt usage. These factors have a higher prevalence for male than female drivers (Hersman and Rosekind, 2017).

There are two main approaches to improve road safety and reduce collisions: preventing collisions themselves and reducing damage when a crash does occur. The ultimate purpose of safety research accordingly is to find countermeasure strategies that effectively implement those approaches (Loo and Anderson, 2015).

In Kansas, the Kansas Department of Transportation (KDOT, 2015a) has long recognized that teen drivers are a group meriting particular focus to reduce their crashes. While teen drivers ages 14 to 19 are one of the foci of KDOT's Strategic Highway Safety Plan (SHSP), identifying strategies to lessen the incidence of crashes involving teens, along with the means of implementing them, is "Complicated by the fact that, besides simple inexperience, other characteristics tend to set teens apart from other drivers" (KDOT, 2015a). After several years of improving metrics, it appears that overall teen crashes have begun to increase in the past few years. The annual number of crashes (of all types) involving teen drivers in Kansas increased by 644 (from 10,715 to 11,356) from 2013 to 2016. This concerning trend shows a need to identify the associated factors and the hazardous locations for this group in order to better target safety improvements.

From this standpoint, there is a need for investigating the location of traffic crashes that involve teen drivers. Using Geographic Information System (GIS) is one of the best options to investigate spatial data. Among the wide range of GIS applications, the field of transportation has received attention, and a specific branch of GIS called Geographic Information Systems for Transportation (GIS-T) had been developed. GIS-T refers to the principles and applications of applying GIS technologies to transportation problems (Rodrigue, 2017). Therefore, GIS-T could be a method to improve road safety by analyzing geospatial data of crashes and identifying patterns of those crashes and their associated factors. GIS-T provides a visual representation of collision locations, which maps clustering of collisions based on selected parameters related to drivers, vehicles, environment, or land use.

A large number of transportation agencies across the US conduct spatial analysis to characterize traffic crash locations and to analyze relationships using different spatial statistical tools in GIS software such as ArcGIS. For example, the Spatial Statistics toolbox in ArcGIS can be used to identify high-crash roadway locations, to conduct the nearest neighborhood analysis, to investigate the existence of clustering (e.g., hotspots), and to model the spatial relationships. With these statistical outcomes, state and local transportation safety officials will be able to identify hazardous locations better, manage and control improvements to enhance safety for teen drivers, and provide a baseline to compare any changes in the future.

PROBLEM STATEMENT

Finding countermeasure strategies that effectively reduce the number and severity of crashes represents the primary goal of traffic safety researchers. Most of the existing studies investigate the effects of demographic differences and nonspatial factors of traffic crashes such as gender, age, DUI, the presence of passengers, distractions, road conditions, and other variables (Oris, 2011). Besides these factors, it is necessary to integrate all the various factors and discover other potential factors thorough understanding of how crashes are related and patterned spatially. This research integrates and analyzes spatial and nonspatial data of crashes involving teen drivers aged 14 to 19 using ArcGIS through mapping crash locations and visualizing their patterns spatially and statistically.

In the US, in general, and in Kansas, specifically, adopting the spatial analysis methodology to identify the hazardous locations (hotspots) for teen drivers and factors behind their crashes is underutilized. Even though Kansas established a comprehensive Strategic Highway Safety Plan (SHSP) in 2015 in order to reduce the number of traffic injuries and fatalities, its goals were not met because of a prioritization strategies shortage (KDOT, 2015a). Using the spatial analysis technique herein will help planners, police, and the public to understand associated teen driver factors and to visualize locations requiring traffic safety improvements. They will receive practical tools that enable them to understand teen-related crashes better and provide a safer environment for teen drivers, especially at hazardous locations.

RESEARCH OBJECTIVE

The primary objectives of this study include conducting a descriptive analysis and a statistical geospatial analysis of teen-related crashes to identify spatial patterns and hotspots for crashes involving teen drivers aged 14 to 19. Also this research focused on analyzing and modeling spatial factors that contribute to changes in the number of crashes and crash patterns using ArcGIS Pro (version 2.3). Finally, a model was developed that could predict future number of crashes at either the county or state level.

The particular objectives of this research were to:

- Identify crash frequencies of teen-related crashes by three factors: crash types and conditions, driver characteristics, and vehicle types regarding crash locations. Determine the location of the highest number of crashes in terms of crash types such as the number of vehicles involved, presence of other factors (pedestrian, cyclists, animals, and/or fixed objects), type of impacts (rollover, rear-end, head on, side-impact, and sideswipe). Furthermore, the crash conditions with regard to the severity of the crash, road conditions, weather conditions, light conditions, and time and date of crashes were identified. The identification of the most hazardous locations in the descriptive analysis phase supported the determination of study areas for the spatial analysis phase.

- Identify spatial patterns and hotspots for crashes involving teen drivers in Kansas and test the hypothesis that there are statistically significant spatial patterns for the crashes, as well as that the crashes tend to cluster within the study area.
- Model spatial characteristics associated with the crash patterns to identify the factors that significantly influence the number of crashes of teen drivers.

The procedures that followed in conducting this research are illustrated in Figure 1 and will be discussed in detail in the following chapters. This research represents an effort to improve highway safety by demonstrating how crashes can be better identified on the statewide level or for smaller geographic regions using spatial analysis techniques. This research helps traffic safety–related parties such as Departments of Transportation and local agencies to identify their emphasis areas and prioritize those locations or factors for the implementation of safety improvement processes.

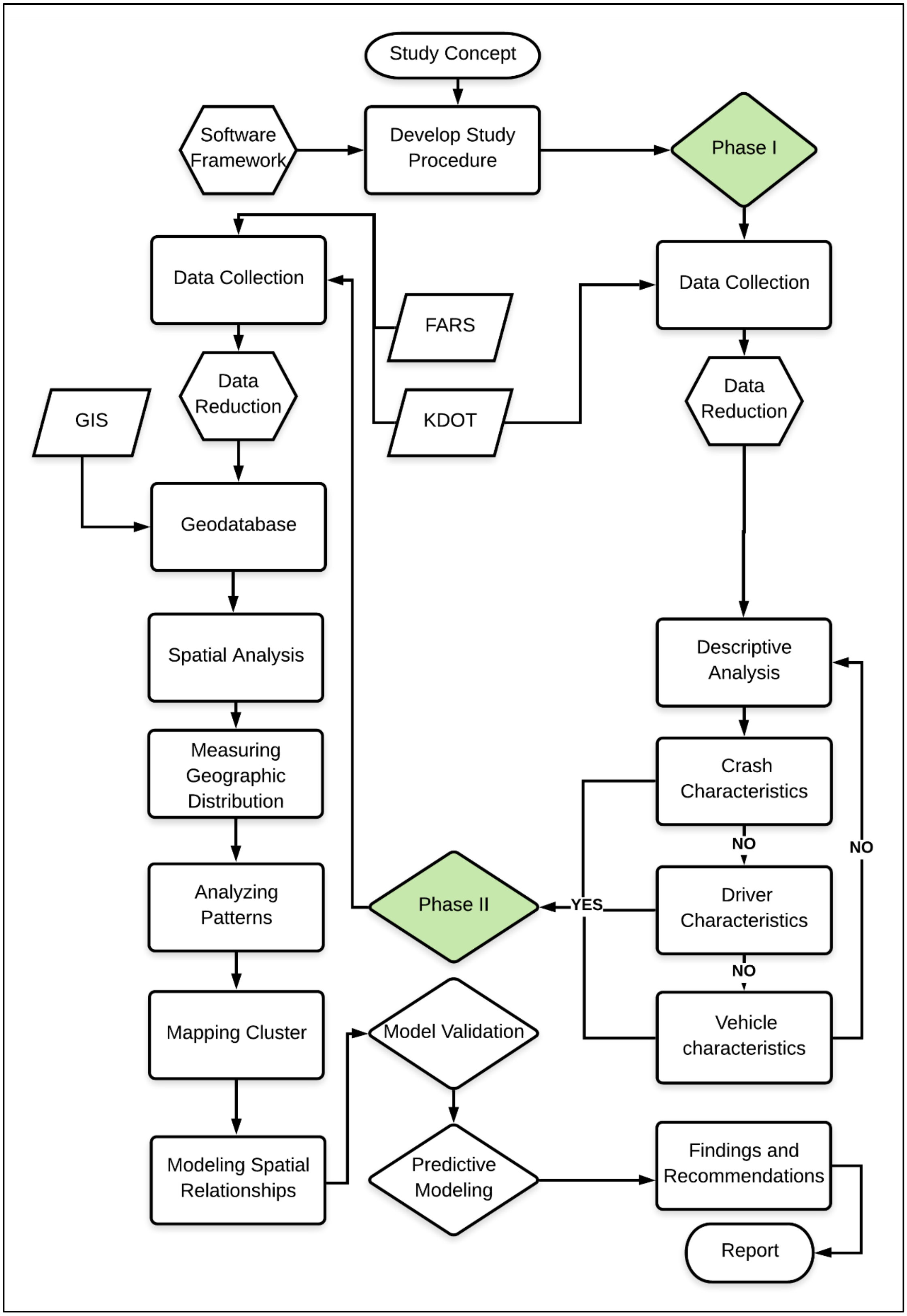


Figure 1. Flowchart of Research Plan

ORGANIZATION OF DISSERTATION

This dissertation is structured into six chapters: Chapter I presents an introduction to traffic safety conditions globally and nationwide regarding teen drivers and approaches to improve their safety, and discusses the significant goals of conducting this research. Chapter II provides a review of the related literature, including studies of teen drivers' safety using various statistical and spatial analysis techniques. Chapter III describes the approaches used to conduct the research. The data collection and data reduction process to create a database for analysis and provides a background on the study area are illustrated in Chapter IV. The descriptive analysis, spatial analysis techniques, testing hypothesis, and the results of the conducted analysis are presented in Chapter V. In the last chapter, Chapter VI, the findings of the results and recommendations for extending this research are discussed.

CHAPTER II. LITERATURE REVIEW

Traffic safety improvements require identification of hazardous locations and causality through the crash analysis process. Different techniques have been used for crash analysis, and each of those methods have specific objectives and have their pros and cons. In general, crash analysis techniques could be classified into two groups: descriptive analysis techniques and predictive analysis techniques.

DESCRIPTIVE ANALYSIS

This type of analysis summarizes and analyzes the crash data to help interested parties pinpoint contributing factors and hazardous locations and prioritizing them for safety improvements. Various factors contribute to traffic crashes, but human errors are present in 95 percent of them (Herbel et al., 2010). These errors cannot be eliminated entirely, but their numbers and outcomes can be reduced by considering the contributing crash factors. Crash factors may be related to roadway factors, such as roads geometry and standard; human factors, such as drivers, and road user behavior; vehicle factors which contribute to crash avoidance and crash protection; and environmental conditions such as rain, snow, ice, and wind (Herbel et al., 2010). To understand the factors that contribute and to reduce the number and severity of crashes, William Haddon combined a timeframe of crashes with crash factors in the "Haddon Matrix."

The Haddon Matrix is a table showing the human, vehicle, and environmental factors that can interact during a crash, against the time sequence of the crash divided into three phases (pre-crash, crash, and postcrash), as shown in Table 1 (O'Neill and Mohan, 2002). For example, apart from providing a crossing guard and telling children not to run across the roads on the way to school (pre-event phase), we can determine a school zone to reduce the speed limit (pre-event phase) thus if a child does get hit, the level of severity will be less severe (Peden et al., 2004; YOURS, 2012). The main approaches for safety research revolving improving teen drivers' safety were the first two phases of the Haddon Matrix, preventing collisions, and reducing the severity of crashes. To understand why young drivers are more at risk of traffic crashes, the next sections review literature related to the contributed crash factors for this age group

Table 1. Haddon Matrix for Traffic Crashes, adapted from (IPP, 2009)

Phases	Factors			
	Human	Vehicle	Physical Environment	Social Environment
Pre-event (crash prevention)	• Driver vision • Alcohol impairment • Driver experience/ ability	• Maintenance of brakes, tires • Speed of travel • Load characteristics	• Adequate roadway markings • Divided highways • Roadway lighting • Hazardous intersections • Road curvature • Adequate roadway shoulders	• Public attitudes on drinking and driving • Impaired driving laws • Graduated licensing laws • Speed limits • Support injury prevention
Event (Injury prevention during the crash)	• Spread out energy in time and space with seat belt and/ or airbag use • Child restraint use	• Vehicle size • Crashworthiness of vehicle— "crush space", integrity of passenger compartment, overall safety rating • Padded dashboards, steering wheels, etc.	• Guard rails, median barriers • Presence of fixed objects near roadway • Roadside embankments	• Adequate seat belt and child restraint laws • Enforcement of occupant restraint laws • Motorcycle helmet laws
Post-event (Life sustaining)	• Crash victim's general health status • Age of victims	• Gas tanks designed to maintain integrity during a crash to minimize fires	• Availability of effective EMS systems • Distance to quality trauma care • Rehabilitation programs in place	• Public support for trauma care and rehabilitation • Emergency Medical Services (EMS) training

HUMAN FACTORS

To understand why teen crash risk is higher than other age groups and why some subsets of teens have more crashes by comparison, it requires an understanding of teen driving and contributing factors that influencing their driving. Shope and Bingham (2008) categorized these factors, as shown in Figure 2. Some of these factors affect all teen drivers such as driving experience, while others could affect a subset of teen drivers, which increase the probability of crashes, such as the tendency for sensation seeking.

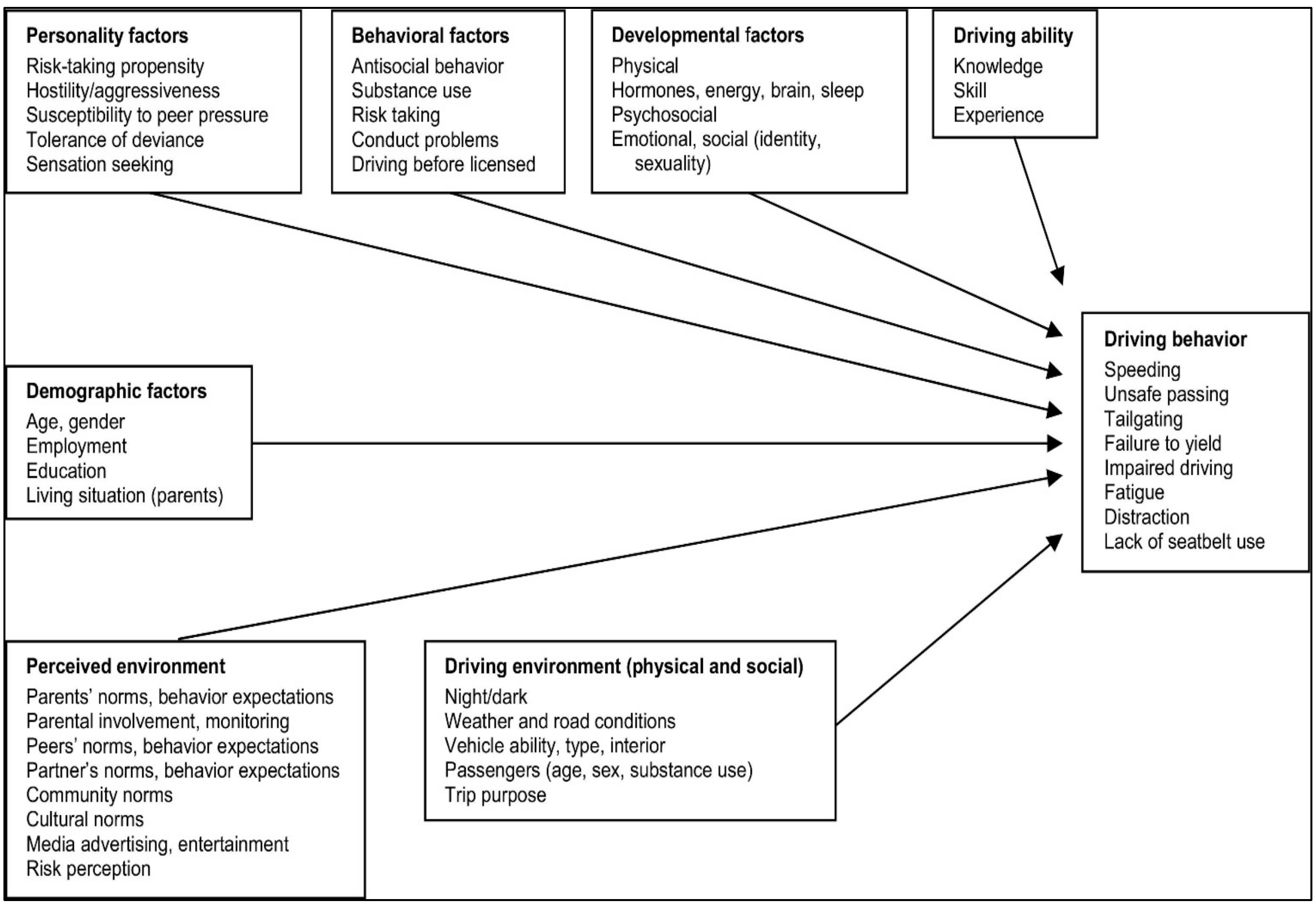


Figure 2. Factors that affect teen driving behavior (Shope and Bingham, 2008)

Curry et al., (2011) examined crashes involving teen driver errors based on categories defined by The National Highway Traffic Safety Administration (NHTSA). The categories are (Singh, 2018):

- Recognition errors, such as inattention and distraction;
- Decision errors, such as driving too fast for the conditions and aggressive driving behavior;
- Performance errors, such as improper steering; and
- Non-performance errors such as falling asleep.

The researchers concluded that of 822 teen drivers aged 15-18 years who were involved in 795 serious crashes, driver errors were a critical reason for 95.6 percent of them in comparison to vehicle and environmental factors (Curry et al., 2011), whereas this ratio was 87.7 percent for all age drivers (Dingus et al., 2016). Among those crashes, recognition errors accounted for 46.3 percent, decision error for 40.1 percent, and performance errors for 8 percent. Also, the gender of teen drivers had no significant impacts on classified errors (Curry et al., 2011).

Experience and Age

During 2006-2015, the proportion of licensed teenage drivers declined and has not rebounded (Shults et al., 2015). The decline was by nine percent and the proportion of those who did not drive during an average week increased by 8 percent (Shults and Williams, 2017). Although teen drivers drive less than other age groups, as shown in Figure 3, they have the highest crash rate (Block and Walker, 2008). Additionally, the number of licensed driver fatalities per million licensed drivers and billion kilometers of travel showed that teen drivers and older drivers represent the highest fatality rates in comparison to middle-aged drivers, as shown in Figure 4

(Evans, 2004). The per-mile crash rate of 16-year-old novice drivers is approximately ten times higher than of adult drivers (McKnight and McKnight, 2003). Even in the same age group, the crash rate per mile driven is three times higher for drivers that are 16 to 17 years old as compared to older teen drivers that are 18 to 19 years old (CDC, 2017).

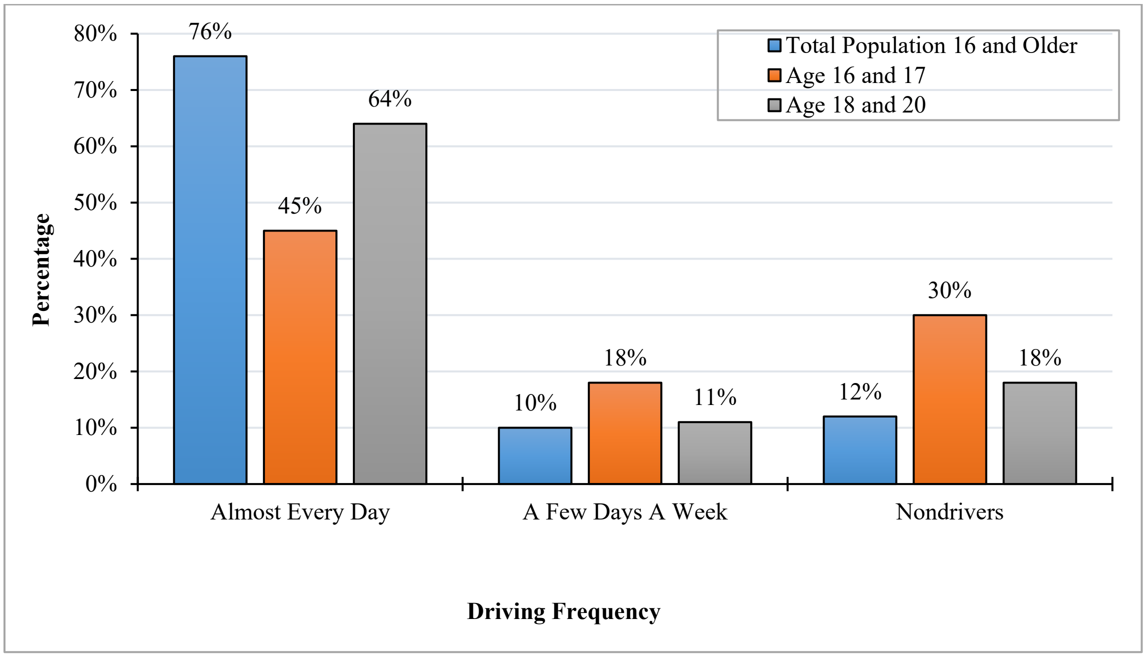


Figure 3. Driving Frequency: Youth vs. Population, adapted from (Block and Walker, 2008)

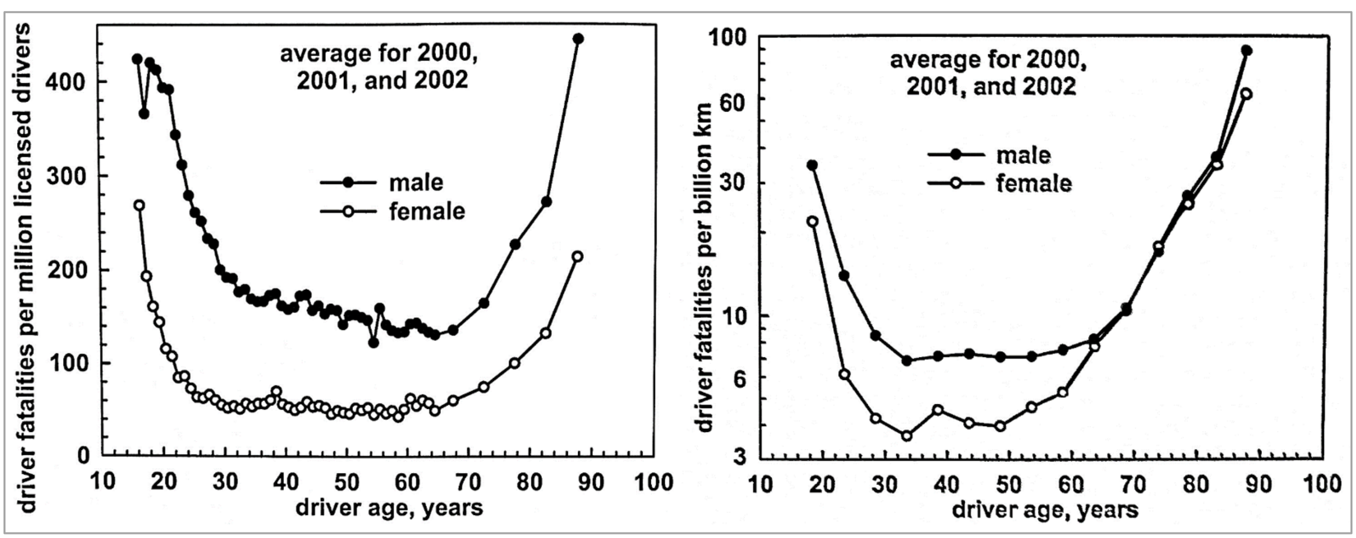


Figure 4. The License and Distance Rate Versus Gender and Age (Evans, 2004)

Peek-Asa et al. (2010) compared, based on crashes occurred on rural and urban roads, crashes involved younger teen driver, aged 10-15 years old, with teen drivers aged 16-18 in Iowa. Even though drivers are not legally allowed to drive until 14, Iowa crash database for ten years (1995-2004) included a sufficient number of unlicensed drivers under the age of 15. The results of data analysis using logistic regression showed that the crash rates were higher for younger teen drivers in rural areas. The probability of rural teen crashes resulting in fatalities, or severe injuries was five times more than urban teen crashes, and rural crashes were more likely to be associated with crash characteristics such as single-vehicle crashes, crashes late at night, crashes involving failure to yield the right-of-way, and/or crossing the centerline of the road (Peek-Asa et al., 2010).

The frequency of driving and the number of miles driven do not increase linearly with crash frequencies and the most plausible explanation for this is that each time a novice driver drives they have the benefit of the previous experience thereby quickly reducing their crash risk (Shinar, 2007). This indicates that the prime crash causes for this age group may include not only a lack of maturity, which is related to carelessness but also a general lack of experience. Comparisons concerning age and experience indicated that the age of newly–licensed young drivers did not affect the rate of safety-relevant events, while experience did (McGehee et al., 2013). Therefore, age and experience seem to have independent effects on driving behavior and driving safety, and both need to be considered in any attempt to promote safety (Shinar, 2017) and this statement can be broadly generalized for both male and female young drivers (McKnight and McKnight, 2003).

Cestac et al., (2011) found that the young drivers who were licensed for less than one year were affected by sensation seeking[1] more than those licensed over one year. The researchers concluded that the normative influence was stronger for young drivers who had been licensed between one and three years and the perceived behavioral control increased as their driving experience increased (Cestac et al., 2011). Even though young drivers are inexperienced, they tend to be adventuresome and can be more likely to drive during high-risk conditions such as at night, when they are fatigued, or when they are under the influence of drugs or alcohol (Shinar, 2017). Consequently, they are over-represented in single-vehicle crashes, which are the type of crashes that most closely associated with the risky driving environment and situations such as driving at night, speeding, fatigue, and driving under the influence of alcohol or drugs (OECD, 2006).

The most commonly cited reason given in single-vehicle crashes (80 percent) was a decision error (Shinar, 2017). McDonald et al. (2014) found that teen drivers were more likely to be involved in crashes due to decision errors than other age groups. According to Singh, decision errors comprised of: driving too fast for conditions or on curves, erroneous assumption of others' actions, illegal maneuvers, and miscalculation of a gap or another driver's speed, which accounted for about 33 percent of drivers' fault that caused crashes (Singh, 2015).

Risky Behaviors

Traffic injuries among youth is a serious public health problem in many countries, and it is higher in low-income and middle-income countries than high-income countries (WHO, 2007), because young drivers in those countries tend to drive in more risky conditions. For instance,

---

[1] Sensation seeking is a predictor of reckless driving.

young drivers in Colombia, a middle-income country, were found to be considerably more risky as drivers than young drivers in Australia, a high-income country (Scott-Parker and Oviedo-Trespalacios, 2017).

In France, Delhomme et al. (2012) tested the hypothesis that speeding corresponds more closely to the definition of sensation seeking than anger on risky driving behavior. The results of their regression analysis conducted on 143 young drivers concluded that sensation seeking was a better predictor of speeding than anger (Delhomme et al., 2012). Teen drivers pose a risk not only to themselves and other road users but also to their passengers. McDonald et al. (2015) evaluated the injury and fatality rate of different modes going back and forth to schools in North Carolina. They found that riding with teen drivers was the most dangerous mode and had higher injuries and fatalities in comparison to riding school buses, riding with adults, biking, or walking. Restricting teen drivers, through a Graduated Driver Licensing (GDL) process, from driving with passengers under 21 years old at drop off and pick up school times could reduce this type of crashes (McDonald et al., 2015).

Fatigue and Drowsy Driving

Fatigue and drowsiness are factors in many traffic crashes and fatalities. Young drivers are overrepresented in distracted and drowsy driving crashes in comparison to other age groups (Hersman and Rosekind, 2017). In respect of drowsy driving and high school start times, Vorona et al. (2011) tested the hypothesis that early high school start times contribute to a higher crash rate in comparison to delaying high school start times. The researchers compared two high schools in two adjusted and demographically similar cities, Virginia Beach and Chesapeake, Virginia. Two years of crash data involving teen drivers aged 16-18 years old were analyzed, and results indicated a significant increase of crash rate in Virginia Beach, where high schools start 75-80 minutes earlier (Vorona et al., 2011).

In Fairfax County, Virginia, where the high school days begin before 7:30 a.m., Hellinga et al. (2007) analyzed 4,225 crashes involving drivers aged 16-17, which occurred between 2001 and 2004 to better understand the temporal pattern of teen driver crashes during school commute times (6:30-7:30 a.m. and 2:00-3:00 p.m.). The results indicated that during weekday school commute times, teen driver crash involvement spiked and the crashes were more likely to involve multiple vehicles and more than one teen driver but less severe in comparison to other times (Hellinga, McCartt, and Mandavilli, 2007).

Another study was funded by the National Highway Traffic Safety Administration (NHTSA) to test the hypothesis that crash rates could be reduced due to later school start times. The analysis of time-series data involving 16-17 years old drivers involved in crashes during the school day indicated that shifting school start time by 75 minutes in Forsyth County, North Carolina decreased number of crashes among this age group (NHTSA, 2015; Foss et al., 2015). The research supported a previous study, which indicated that later school start times might decrease teen drivers' risk of crashes (Danner and Phillips, 2008).

The shifting in school start time was found to be beneficial to improve not only academic performance of teens but also reaction time, hazard perception, alertness, and decision-making ability and thereby traffic safety (NHTSA, 2015). The school start time shifting to 8:30 a.m. or later reduced teen driver exposure on school days. As Hakkert and Braimaister (2002) stated through the equation: "$\text{SAFETY}_{\text{(severity)}} = \text{RISK}_{\text{(trend)}} * \text{EXPOSURE}_{\text{(trend)}}$," reducing exposure improves traffic safety. To clarify the impact of exposure on teen drivers' safety, the study conducted by Stone and Runyan (2005) represents a supportive example. The researchers compared crashes involving teen drivers in two counties in North Carolina having open-lunch policies, which enabled teens to drive off campus for lunch, with another county without this lunch policy. The results showed that the crash rates in the counties with open-lunch policies were significantly higher for teen drivers over lunch hours (Stone and Runyan, 2005).

Distraction and Crash Types

For safe driving, drivers need to avoid the three modes of distractions: visual, manual, and cognitive (DMV, 2018b). Visual distractions are activities that divert drivers' eyes off the road such as looking at their cell phones, while manual distractions could be any physical activities that make drivers take one or both hands off the wheel such as eating or grooming. Cognitive distractions include activities that pull away drivers' focus from the driving task such as talking to a passenger or on their phone (NHTSA, 2018a). To demonstrate the relationships between driver distraction and safety, researchers analyzed data gathered from various sources such as in-vehicle data recorder, police reports, and surveying.

Carney et al. (2016) analyzed 400 teen driver rear-end crashes captured by in-vehicle event recorders, and they found that over 75 percent of the crashes were due to the driver's involvement in a distracting behavior with no significant differences by gender. Compared to solo driving, the presence of at least one passenger (parents or peers) with teen drivers affect drivers behavior and increases the crash risk (Ouimet et al., 2015; Ouimet et al., 2013; Carter et al., 2014; Foss and Goodwin, 2014; Curry et al., 2012). Ouimet et al. (2015) reviewed seven previous studies that examined the effects of the presence of passengers on teen driver performance. It was found that the fatal crash risk for teen drivers with at least one passenger was between 1.24 to 1.89 times higher compared to when they has no passengers. Most of these studies made no distinction as to the age of the passengers, but one study by Chen et al. (2000)

found that when the passenger(s) is (are) at least 30 years old, then no significant increase in fatality rate was found. This seems to indicate that there is a performance difference when driving with an adult (presumably parents) compared with driving with other teens.

Ouimet et al. (2015) also reviewed ten previous studies examining nonfatal and combined fatal/nonfatal crashes, but the results were mixed with significantly increased risk and non-significant associations. Ouimet et al. (2015) concluded from studies that examined fatal crashes; there was an increased risk for drivers with two passengers and three or more passengers by 1.7 to 2.92 times higher compared with solo driving. Williams and Tefft (2014) found that of 16- and 17-year old drivers involved in fatal crashes, 57 percent had at least one passenger. Therefore, strengthening restrictions of passengers could reduce the number of such crashes.

The type and level of severity of crashes involving teen driver often changes with the kind of distraction. For instance, at intersections, the teen drives distracted by passengers or cognitively distracted teen drivers are more likely to be involved in rear-end or regular crashes than fixed-object crashes (Neyens and Boyle, 2007). Neyens and Boyle (2008) concluded that the rear-end crashes involving teen drivers were more severe when teen drivers were distracted by cell phones or by passengers than when they were distracted by in-vehicle devices or by lack of attention. In spite of that, young drivers are more willing, in comparison to other age groups, to answer or initiate a call regardless of the road and traffic conditions and they had poorer performance on the driving task when engaged in calls (Tractinsky et al., 2013).

Regarding crash types and associated factors with crashes involving teen drivers, Braitman et al. (2008) analyzed data gathered from interviewing 260 16-year-old teen drivers in Connecticut, who were involved in nonfatal crashes. Results from a chi-square statistical test indicated that 76 percent of teens were reported at-fault had at least two common factors (speeding, slippery road, or loss of control) associated with more than half of at-fault crashes, which encompassed all the three common types of crashes, rear-end, run-off, and violation of right-of-way (Braitman et al., 2008). In overall crash types, the gender of the teens had no significant differences, but in run-off-road crashes, males were more likely to be involved than females by 12 percent (Braitman et al., 2008).

A swift engagement of technology and its rapid evaluation have tremendous impacts, positively and negatively, on drivers' safety, especially young drivers. Mobile devices present a major distraction source for young drivers who already lack driving experience. Engaging in texting and dialing cell phones increase the risk of crashes for teen drivers more than other age groups (Klauer et al., 2014). On the other hand, technology provides different in-vehicle tools to support safe driving such as Intelligent Speed Adaptation (ISA), the Electronic Stability Program (Scott-Parker and Oviedo-Trespalacios, 2017), collision avoidance systems, and lane departure warning systems (Lee, 2007). The next few years will bring more distraction devices and more complex equipped vehicles.

Some people might ask, do teen drivers know which activities are considered distractions? To answer this question, Westlake and Boyle (2012) analyzed a data set gathered from surveying 1893 teen drivers in Iowa, and they found that 80 percent identified texting as a distraction and risky behavior. Although the majority of them recognized the distraction activities, a subgroup of teens always engaged in distracting activities (Westlake and Boyle, 2012). Any activity that makes teen drivers glance away from the forward road because of involvement in secondary tasks, regardless of those secondary tasks, increases the probability of crashes (Simons-Morton et al., 2014).

Hazard Perception (HP)

Hazard perception is the ability to anticipate dangerous traffic situations (Sagberg and Bjørnskau, 2006). Hazard perception changes with the amount of driving experience and with the type of hazards that drivers face. This ability is faster for experienced compared to inexperienced drivers due to faster processing in detecting, evaluating, and responding to risks (Huestegge et al., 2010; Crundall et al., 2012). To test the hazard perception of novice drivers before permitting them to drive, many countries, such as the UK and Australia, integrated the Hazard Perception Test (HPT) into their GDL program. The HPT is a computer-based test in the training and licensing process that assesses drivers' ability through video scenes to identify dangerous traffic situations and respond appropriately (Wetton et al., 2011). Scialfa et al. (2011) developed and assessed a new version of HPT for Canada to improve the ability of novice drivers to identify hazards quickly and correctly on the road. The outcome of the study was a set of 18 scenes with 15 minutes of testing time that could characterize drivers from novice to experienced (Scialfa et al., 2011).

Gender

Male teen drivers generally are involved in more risky driving behaviors than female teen drivers (Rhodes and Pivik, 2011). Therefore, young male drivers are particularly at risk with death rates of up to three times those of young female drivers (OECD, 2006). This is mostly because of male young drivers' intention to speed as an impact of sensation seeking, and injunctive norms more than young female drivers (Cestac et al., 2011). Because male teen drivers have a higher exposure to risky driving and thereby a higher chance for being in traffic crashes, Taubman–Ben-Ari et al. (2015) analyzed a sample of 121 young male drivers (17-21.5 years) to evaluate parents' contributions to their newly-licensed teen drivers in Israel using Poisson-based models.

The analyzed data were gathered using in-vehicle data recorders and self-report questionnaires completed by the young drivers. The results indicated that parents-especially fathers- represented a vital model and were positively associated with higher risk driving and increased distraction of their teen drivers (Taubman–Ben-Ari et al., 2015; Bingham et al., 2015).

Family climate and parents play a critical role in their young drivers in regard to safe driving (Taubman–Ben-Ari, 2014), especially in the learning period. Jewett et al. (2016) analyzed a subset of data in Washington DC, which included 456 parents of learners and newly licensed teens ages 15-18. The resultant log-linear regression model showed that 80 percent of parents had restrictions for their teen drivers regarding drinking and driving, seatbelt usage, and using the cell phone, but only nine percent of these restrictions were formalized as a parent-teen driving agreement. Teens quickly develop their basic vehicle handling skills, and that makes parents mistakenly believe that their teen is ready to drive independently. Therefore, the results indicated that 61 percent of teens' parents worry "a lot" for their teen's safety in comparison to 36 percent of newly licensed teens' parents when they were provided with four levels of worry intensity ("a lot" versus "somewhat," "not very much," or "not at all") (Jewett et al., 2016).

VEHICLE

The type of vehicle that teenagers drive during their initial months of licensure is one of the critical decisions that parents make regarding their teens' safety. In this respect, Cammisa et al. (1999) conducted a longitudinal study on newly-licensed teen drivers and their parents in four states, Connecticut, Delaware, New York, and New Jersey, to investigate the process of vehicle selection for teens during learning and once they were licensed. The study concluded that teen drivers were more likely to drive old vehicles, which means vehicles lacking modern crash protection features, and small vehicles, which represents lower crash protection in comparison to large vehicles. In term of crash rates, teenagers who owned vehicles drove more and had a higher crash rate than nonowners (Cammisa et al., 1999).

Hellinga et al. (2007) interviewed 300 parents of 16-17 years newly licensed teen drivers from Minnesota, North Carolina, and Rhode Island examining parental decisions and knowledge of vehicles driven by teenagers as to safety. The results indicated that even though most of the parents realized the importance of major safety criteria in the vehicle, they chose vehicles with inferior crash protection quality or unsafe vehicles for newly-licensed teen drivers. Most of the chosen vehicles were small, pick-ups, small SUVs, or sports cars instead of larger and heavier cars, which are considered safer for all drivers (Evans, 2004). Those choices were mostly decided based on economic issues (purchase price, maintenance cost, insurance cost, and gas consumption) not safety features (airbags and brake systems) (Hellinga, McCartt, and Haire, 2007). These finding supported the results of a study conducted by Williams et al. (2006) through interviews of 3,500 teenagers and their parents in Connecticut. The type and size of the chosen vehicles were not ideal for newly-licensed teen drivers, and most of them did not contain key safety features such as electronic stability control (Eichelberger et al., 2015).

Several studies investigated the relationship between vehicle model year and fatal crash rates. The results indicated that the age of vehicles involved in crashes is correlated with the severity of the crashes and newer vehicles are safer and had lower fatal crashes than older model year vehicles (Ryb et al., 2011; NHTSA, 2013, 2018c; Glassbrenner, 2012). Age of the selected vehicles for teen drivers was also studied, by Williams et al. (2006) who studied types of vehicles driven by young beginners and concluded that the majority of teen drivers at licensure and 12 months later were driving older small vehicles. Eichelberger et al. (2015) conducted a study in 2014 and found that 83 percent of teens drove used vehicles (< 2012 model year) when they first began driving.

ENVIRONMENT

Some studies emphasized the difficulty of novice drivers in adaptation to changes in temporal and spatial driving environments, such as speed reduction in the presence of reduced visibility or curved segments. In comparison to experienced drivers, novice drivers had higher hazard response times, greater speed and steering variability, and were the most drivers to have collisions when driving in environments like foggy weather, at horizontal curves, and at intersections (Mueller and Trick, 2012; Li et al., 2015; Borowsky et al., 2010; Scialfa et al., 2011; Borowsky et al., 2009).

As shown in Table 1, public attitudes on drinking and driving is a critical social environment factor that has a significant impact on traffic safety. For instance, in the Mississippi Delta region, where there are high poverty rates, low school graduation rates, and a general lack of resources, Muilenburg et al. (2007) investigated the prevalence of risky behaviors such as alcohol use of pre-driving teens in the rural area. Among 290 middle school students, in the 7^{th} and 8^{th} grade, aged 12-16, about 17 percent reported drinking and driving and 45 percent–with no gender and racial differences– reported riding with a drinking driver. The results showed that alcohol-related driving behaviors were not only performed by legally licensed drivers, but also by those illegal young drivers who have access to vehicles (Muilenburg et al., 2007).

Heck and Nathanie (2011) examined differences in driving behaviors between teens who live in an urban area compared with those living in suburbs and rural areas. The results from a survey included 1940 high school seniors at twelve high schools in California showed that teens from urban areas were significantly less likely to be licensed and they spent less time behind the wheel. On the other hand, teens from suburban and rural areas were more likely to be associated with unsafe driving behaviors such as reckless driving and drinking and driving (Heck and Nathaniel, 2011). Regarding the variety of driving frequency among high scholar teens aged 16 or older, Shults et al. (2015) concluded from 42 states that the ratio of teens who drove was between 53.8 percent and 90.2 percent.

Driving prevalence depends on where young drivers live. For instance, teens in Midwestern and mountain states drove more than other states (Shults et al., 2015). Another example might be driving in states with large rural road networks. Crashes occurring on rural roads are generally more severe than those on urban roads due to the uniqueness of driving conditions on rural roads such as higher speed limits, different geometry, longer trips, less enforcement, less controlled in adverse weather conditions (Kumfer et al., 2017; Peek-Asa et al., 2010). To improve the knowledge of rural road safety for teen drivers, Kumfer et al. (2017) suggested adding a flash-based computer education tool as an intervention tool in a phase of GDL programs.

Graduated Driver Licensing

The crash rate of teen drivers, especially 16-17 years old, is higher than all other age groups. In order to address this threat in the United States and Canada, the GDL system was introduced in the mid-1990s and has been enacted into law by all US states replacing previous laws that allowed immediate full access to driving (Williams and Shults, 2010). The first comprehensive GDL system was applied in Florida in 1996 (Preusser and Tison, 2007). In the GDL system, novice drivers have an opportunity to learn practically basic driving skills under fully licensed drivers' supervision (typically parents) in the "learner-licensed" level, which allows teenagers to gain a significant amount of driving experience. The system during the "intermediate" level allows a learning driver to drive independently and imposes restrictions on learners to limit their exposure to high-risk driving conditions such as driving late at night and/or driving with teenage passengers (Foss and Williams, 2015).

The GDL programs with restrictions on both driving at night and driving with teenage passengers had substantial impacts on reducing the number of fatal crashes by 23 percent for 16 year–old drivers (Masten et al., 2011; Preusser and Tison, 2007; McCartt et al., 2010). Baker et al. (2006) concluded that comprehensive GDL programs contributed to reducing about 20 percent of fatal crashes involving 16 year–old drivers through analyzing 11 years of crash data (1994-2004) for 43 states in the United States. The most effective program in lowering fatal crash rates for 16 year–old drivers included: restrictions on age, nighttime driving, passenger restrictions, at least 30 hours supervised driving, and at least a three–month waiting prior to moving to the intermediate level (Baker et al., 2006).

The duration of each driving license level is different among the states. There is no standard number of months for the learner license level or a number of supervised hours for the intermediate level. Ehsani et al. (2013a) determined the effects of the learning and supervising periods on fatal crashes involving teen drivers by comparing the states that mandated the learner's license to be six months and the states that mandated three months. They concluded that in order to achieve a significant reduction in fatal crash rates involving teen drivers, the learning duration needs to be at least six months long. Senserrick and Williams (2015) stated based on literature that the range of supervised hours (80–120 hours) was effective in reducing fatal crash rates.

The risk patterns among this age group varied by driving license status, low in the learning period, but high after licensure, late at night, and if passengers present. These variations customize the basis for GDL programs, which were designed to promote low-risk driving and to curb high-risk driving (Williams, 2003). The number of teenagers who did not get a license before their 18th birthday most likely are from low-income families, who did not participate in the GDL program due to the cost of the program or cars (Shults et al., 2015; Tefft et al., 2014). In this regard, six states and the District of Columbia extended some GDL restrictions such as driving at night, driving with passengers, and the learning period to all who apply for driver licenses at age 18 or older.

For example, New Jersey added all GDL rules including night curfew and passenger restriction to new 18-20 years old applicants and added a minimum three months to the learner licensed level and one year to the intermediate licensed level for 21 or older applicants. Curry et al. (2017) examined this step by analyzing crash data of more than one million novice drivers, who obtained their intermediate license at different ages between 2006 and 2014, in New Jersey. The results supported New Jersey's modifications on GDL rules for 17-20 years novice drivers, but found no compelling evidence for restrictions regarding novice drivers aged 21 and older. In the same context, many states made various adjustments to their GDL system in order to achieve further reduction of crashes involving teen drivers as New Jersey did. For instance, Massachusetts built on its first GDL law implemented in 1998 to enact its second GDL system in 2007. The new GDL was associated with decreasing crash rates, and fatal crashes involving the novice driver population (DePesa et al., 2017; Kaafarani et al., 2015).

In Nebraska, teens have two options to apply for an intermediate-level provisional operators' permit: completing the education safety course offered by the Department of Motor Vehicles (DMV) and passing required tests or submitting a 50-hour certification form signed by an adult (usually parents). Shell et al. (2015) compared both options in term of crashes and traffic violations through an examination of eight years of data (2003-2010). The results indicated that teens who took the driver education course had fewer crashes, injuries or fatalities, traffic violations, and alcohol-related violations during the first two years of their intermediate stage, which means that driver education courses have a positive impact on improving teen drivers' safety on the road.

Shifting to the GDL system had proven its effectiveness through declining crash rates for novice drivers, especially fatal crash types, and it deserves building on it and expanding its effectiveness (Ehsani et al., 2013b; Williams et al., 2016; Williams, 2017; McCartt et al., 2010). However, crash rates of teen drivers are still high compared to other age groups, and some significant factors are still unaddressed by GDL programs (McCartt and Teoh, 2015).

Kansas, as one of many states, have adapted a GDL program, which requires teens to take three steps – mostly age-related – before earning "full-privilege" licenses. The first step is getting the learner's permit when teens turn 14 years old (DMV, 2018a). Once they obtain a learner's permit, they must complete at least 25 hours of behind-the-wheel driving practice supervised by a licensed driver at least 21 years old who holds a valid driver's license before applying for a restricted driver's license (Kansas Department of Revenue, 2016). In the second step, teens a 15 years old can get a provisional license after holding the learner's permit for one year, and completed at least 25 hours of supervised driving (DMV, 2018a). During this step teens must complete the remaining 25 hours of their required 50 hours of supervised driving, plus 10 hours being driven at night, they cannot have any non-sibling passengers, and they cannot use a cell phone (IIHS, 2018). When teens turn 16 years old, they can get a less restricted provisional license. This step reduces some of the restrictions, but they still can only drive to and from work, school, a religious worship service, or a farm-related purpose, and they can have one non-sibling passenger under 18 years old. Once reaching 17 years old and have had the restricted license for six months, teen drivers are qualified for an unrestricted driver's license after passing driving and vision tests (Kansas Department of Revenue, 2016).

Road Geometry

Horizontal curves have been considered a significant safety factor for many years because they are a common factor in many observed traffic crashes, especially the occurrences of roadway departure crashes (Momeni, 2016). To measure the relationships between driving behavior on horizontal curves, Li et al. (2015) conducted a driver simulation experiment on continuous S curves with different driver experiences and gender groups. The results indicated that in comparison to expert drivers, novice drivers were less skilled in controlling vehicles longitudinally and laterally. The novice female drivers were found to have worse performance, in term of speed and lane position within the curve, than other drivers (Li et al., 2015).

New technologies integrated with vehicles, such as collision avoidance systems and lane departure warning systems, might improve teen drivers' safety on the roads, in general, and horizontal carvers, specifically. To examine the benefit of crash warning systems, Jermakian et al. (2017) analyzed data gathered from 40 vehicles equipped with lane departure, potential rear-end, and lane change crash warning systems driven by teen drivers aged 16-17. The results indicated that warning systems improved some teen behaviors such as lane-keeping and turning-signal use, but also had the result of having the drivers follow lead vehicles closer (Jermakian et al., 2017).

Inclement Weather

Inclement weather conditions significantly increased crash and injury rates (Xu et al., 2013). The most inclement weather factors that have negative impacts on teen drivers' safety are heavy rain, bad visibility, strong wind, and snow. Poor decision-making in selecting inappropriate speeds and leaving space around other vehicles for situations like inclement weather conditions, unsuitable visual search strategies and expectation about hazards, and erroneous selective attention are cognition errors that cause many crashes among young novice drivers (Curry et al., 2011).

For instance, Mueller and Trick (2012) investigated the behavior of 19 novice drivers, who were on average 19 years old with an average of six months' driving experience, in the presence of reduced visibility situations (such as fog) in comparison to the behavior of 19 experienced drivers, who were on average 24 years old with an average of eight years driving experience. The analysis of the driving simulator data showed that 25 percent of the young novice drivers were involved in crashes while none of the experienced drivers did. Also, the novice young drivers had greater speed and steering variability, and higher hazard response times in comparison to experienced young drivers.

Night Driving

More than one-third of teen drivers aged 16-17 in the US were involved in fatal crashes at night (9:00 p.m.-6:00 a.m.), and more than 50 percent of them occurred before midnight (Vaca et al., 2006; Shults, 2016). Due to the recognition that driving at night increases the risk of fatal crashes for all drivers, especially teenagers, all states included a night driving restriction in their GDL system. However, the start time of this restriction varies among states. In 24 states, including the District of Colombia, the restriction begins after midnight, and this provides minimal protection for teen drivers because most of their trips end by midnight (Shults, 2016). Therefore, updating the night driving restriction to start at earlier nighttime hours could reduce teenagers' exposure thereby improving their safety on the road.

Teen drivers' risky behaviors are one of the major factors associated with the number of crashes in this age group. To improve the safety behavior of emerging teen cohorts, the service-learning approach can play a decisive role. Goldzweig et al. (2013) conducted a service-learning intervention study in 11 high schools across the United States to improve seat belt use among teens through direct observation techniques and before and after observation. The results from the Mann–Whitney U test to compare collected samples indicated that the intervention had a statistically significant impact in increasing the seat belt use rate by 12.8 percent regardless of race, ethnicity, or gender.

Another type of intervention that may improve teen driver's performance and safety is the video-based feedback. McGehee et al. (2013) examined the effects of a video intervention on reducing safety-relevant events among three groups of 90 newly licensed teen drivers in Iowa: 32 school license holders aged 14.5-15.5, 28 intermediate license holder that had never driven independently, and 30 intermediate license holder with more than four months' driving experience. Half of the participants served as a control group who never got any feedback regarding their driving, while the other half received feedbacks. The results of analyzing the safety-relevant events recorded on event recorders integrated into the participants' vehicles for 24 weeks indicated that the teen drivers that received video-based feedback effectively reduced their rate of unsafe-relevant events in comparison to the control group, regardless of their driving experience and their age (McGehee et al., 2013).

Based on the literature and the evidence presented previously; we can conclude that the high rate of teen crashes does not mean they are bad drivers. It merely means that they are not mature and experienced drivers, and this hinders them in making the appropriate decision in risky driving situations.

SPATIAL ANALYSIS TECHNIQUES

Tobler invoked in his first law of geography that "everything is related to everything else, but near things are more related than distant things."(Tobler, 1970). In terms of traffic safety, it means that there are relationships between crashes that occur in specific locations. These relationships could be the presence of particular trip generation and/or trip attraction centers, driver behavior, physical and social environment conditions, the geometry of roads, traffic-related policies and legislation, etc. Tobler's first law builds a bridge between geography and traffic safety and provides the theoretical foundation for a particular research area, which could be called a spatial analysis of traffic safety. Geography offers exclusive techniques for the research of traffic crashes through mapping and geovisualization, which facilitate studying crashes spatially and identifying their patterns (Hicks, 2009).

Using spatial analysis techniques in conjunction with statistical regression models to determine locations with a high number of crashes enables statistical models to account for the spatial characters of those locations (Loo and Anderson, 2015). Several methods have been developed for crash pattern analysis, such as the network screening, the Kernel Density Estimation (KDE) and the Getis-Ord (G_i*) spatial statistics. The principal objective of these methods is identifying safety hotspots inside roadway networks by comparing existing safety conditions with the expected conditions. When a location has significantly higher crash rates than expected, the area is considered a hotspot, and further analysis could be conducted to identify contributing factors, thereby recommending required actions to reduce the number of crashes.

Kernel Density Estimation

In statistics, KDE is a non-parametric approach to visualize the distribution of data and estimate the probability density function of a random variable (Guidoum, 2013; Duong, 2017). Geographically, the KDE is a geospatial analysis technique that could be applied to point or line datasets integrated with extensive nonspatial attributes (Gibin et al., 2007). The KDE for points is a weighted method that calculates the density of points in a neighborhood around those points, creating a smoothly curved surface that fits over each point (Mitchel, 1999). The value of the created surface is highest at the point and decreases with increasing distance from the starting point until it reaches zero at the search radius (threshold distance) range (Esri, 2018b).

Selecting the value of threshold distance is a challenging process because it has the most significant influence on the smoothness of the resulting density surface. Therefore, extensive researches were conducted which proposed different approaches to estimate an appropriate threshold distance for various applications, not only for traffic crash analysis but also for crime, seismic risk, fire and rescue service, and topography segmentation analysis (Krisp and Špatenková, 2010; Bors and Nasios, 2008; Danese et al., 2008; Nakaya and Yano, 2010; Shafabakhsh et al., 2017). An excessively large threshold distance results in a similar estimated density everywhere, which means close to the average point density, in the study area while an excessively small threshold distance makes the surface pattern focus on the individual point records (Krisp and Špatenková, 2010). The density at each raster cell is computed by summing the kernel surfaces values, where they overlay the cell center, creating a smooth density surface that visualizes hotspots on a map (Hicks, 2009).

The KDE was used for generalization of crash locations for an entire study area and interpreting crash point data in the form of density surfaces. These surface representations provide a realistic continuous model of Hazardous Road Locations (HRL) and show changes in crash density (Loo and Anderson, 2015). Hicks (2009) used the KDE method for temporal and spatial analysis of traffic crashes in Stillwater, Oklahoma. The researcher concluded that the approach represents a useful technique to identify the locations where crashes clustered and to indicate areas where safety improvements were needed. Some research distinguished between KDE analysis of traffic crashes inside one-dimensional linear spaces (road networks) and inside two dimensional geographic spaces by network KDE and planar KDE, respectively (Xie and Yan, 2008).

Anderson (2007) compared KDE, network analysis, and area-wide analysis as alternative spatial statistical methods for identifying road crash hotspots using crash data from 1998 to 2002 in London. The researcher concluded that even though the KDE method was able to identify hotspots quickly and visually from large datasets, the applicability of KDE to identify road crash hotspots was less effective than using it to identify other hotspots such as crime hotspots (Anderson, 2007). This was because traffic crashes are constrained to road and street networks, while other datasets are not so constrained.

Xie and Yan (2008) concluded that using the network KDE method instead of a more generalized KDE is more efficient in analyzing traffic crashes to reveal hotspots when they analyzed crash data from 2005 in Kentucky using ArcGIS. On the other hand, the KDE method determines the spread of risk of crashes in areas around clusters, which identify a higher likelihood for crashes based on spatial dependence. In other words, the KDE seeks to determine risk levels not only at crash points but also in the neighborhood of these points such as schools, shopping centers, and other land uses (Loo and Anderson, 2015).

The KDE techniques were applied widely to identify hazardous locations in roadway networks. Hashimoto et al. (2016) assessed relationships between traffic crashes and city characteristics through developing models for traffic crash density estimation to identify appropriate locations for implementing area-wide traffic calming in Toyota City and Okayama City in Japan. The analyzed data included spatial data of more than 65 thousand traffic crashes that occurred between 1999 and 2010 in these cities. By using KDE, 16 models were developed, and the applicability of using KDE was examined as an explained variable in comparison to that of using raw count data. The results showed a strong positive Spearman correlation coefficient between the predicted number of crashes and the actual number. Namely, there was no significant difference in the correlation coefficient between the developed KDE models and the raw data models (Hashimoto et al., 2016). Shafabakhsh et al. (2017) developed an approach employing the KDE to identify hotspots on urban networks in Mashhad, Iran for three types of crashes: fatal, injury, and Property Damage Only (PDO) crashes. The researchers concluded that KDE provided a significant insight into traffic crash patterns in urban networks, which helps road safety specialists to improve safety.

NETWORK SCREENING

Network screening is a roadway safety management process presented in the Highway Safety Manual for reviewing a transportation network in order to identify crashes pattern, causal factors, and appropriate countermeasures (AASHTO, 2010). This process integrated into six cyclical steps, as shown in Figure 5.

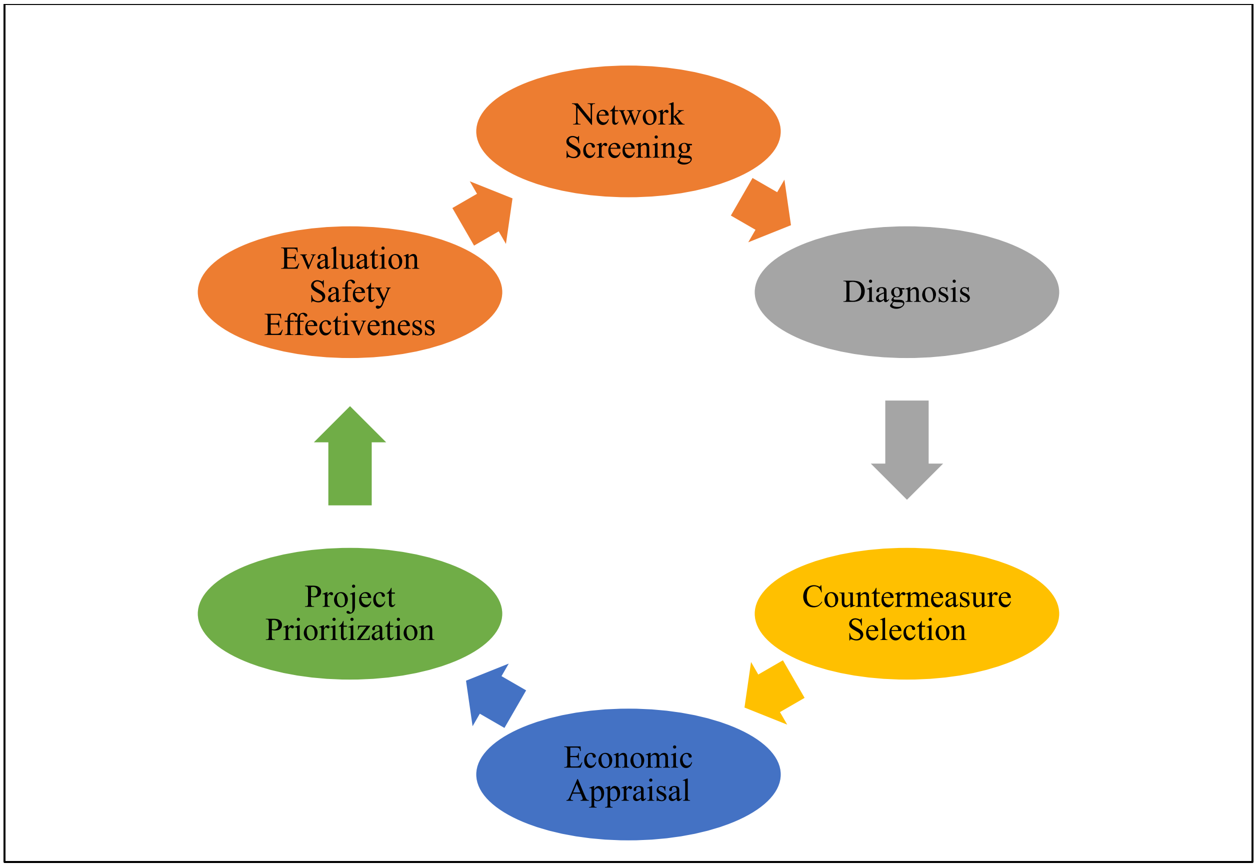


Figure 5. Roadway Safety Management Process Adapted from AASHTO (AASHTO, 2010)

1. Network screening is the first process of analyzing a network to identify the hazardous locations that are expected to get the most benefit from a safety treatment and improvement program across the roadway network (Srinivasan et al., 2016).
2. The diagnosis process identifies the potential safety problems at locations determined in the network-screening step. The method includes reviewing the crash history, traffic operations, geometric characteristics, site conditions, and road user behaviors to produce a list of contributing factors associated with crashes (Gross et al., 2016).

3. Countermeasure selection evaluates available options to mitigate the primary safety issues identified in the diagnosis step. These issues might comprise engineering, enforcement, education, and EMS-related measures (Gross et al., 2016).
4. Economic appraisal compares the relative costs and benefits of the available alternatives using benefit-cost analysis and cost-effectiveness analysis. Crash modification factors (CMFs) are the most common parameter used to assess the safety benefit by estimating the impacts on countermeasures in reducing the number of crashes (AASHTO, 2010).
5. Project prioritization is the process of reviewing possible projects proposed through the previous steps for implementation. This step includes developing an ordered list for selecting projects based on the results of the ranking and optimization processes (AASHTO, 2010).
6. Evaluating safety effectiveness is the final step in the process, which assess the changes in crash numbers and/or severity of crashes due to implemented treatments or projects. This step provides critical feedback on how well allocated funds improved safety and how the results affect future activities and policy revisions (AASHTO, 2010).

The focus herein is the first step of the roadway safety management process, network screening, which identifies hotspots. The HSM classifies network-screening procedures into three major methods based on the analyzed category (segments and nodes). The Sliding Window (SW) method and the Peak Searching (PS) method are used for segments screening and the Simple Ranking (SR) method for nodes screening. Kwon et al. (2013) compared the performance of SW and PS for identifying hotspots on freeway segments against Continuous Risk Profile (CRP) approach. The CRP is a traffic crash data monitoring method to detect hotspots, which do not require segmentation of roadways and its analytical results do not affect the spatial correlation in the crash data (Lam et al., 2009). The results revealed that the false negative rates (i.e., not identifying real hotspots) of SW and PS were comparable and higher than CRP (Grembek et al., 2012; Kwon et al., 2013).

Young and Park (2014) also compared results of identifying crash hotspots obtained from the network screening and the KDE using five years of crash data for the city of Regina, Saskatchewan, Canada. A comparison was conducted regarding the observed severity-weighted crash numbers and the expected severity-weighted crash numbers. The results showed that the KDE method captured a higher number of expected PDO crashes than the network screening method. Because the KDE method takes into account the spatial association with crashes in the entire network and is not limited to single segments, it reduces the effect of regression to the mean (RTM) (Young and Park, 2014).

G_i* SPATIAL STATISTICS

The Getis-Ord (G_i*) spatial statistic method was introduced by Getis and Ord and now is integrated with the ArcGIS software package (Ord and Getis, 1995; Getis and Ord, 1992). The Gi* spatial statistic is a z-score that identifies statistically significant spatial clusters of high values (hotspots) and low values (coldspots) through looking at each feature within the frame of neighboring features in the study area (Esri, 2013; Satria and Castro, 2016). Some important terms in the spatial analysis presented in this definition need to be clarified before proceeding. Statistical significance with spatial tools is mostly based on testing spatial patterns of crashes with the same null hypothesis, which is complete spatial randomness, using p-values and z-scores. This is what differentiates hotspot analysis, which is basically a test for randomness as shown in the definition, from heat maps, which is essentially a choropleth map that shows a density surface based on a selected classification method (Esri Events, 2017a). The other three terms are: feature, neighborhood, and study area. The feature is a polygon that has a value and the value in this research is a count of crashes in each feature, which could be a county, a district, or a Unified School District (USD). The neighborhood is a group of features around any feature while all the features together institute the study area.

The z-scores are standard deviations, and further z-scores from zero are associated with smaller p-values, which are an indication of significant spatial clustering. The Gi* values could be calculated by the equations below:

$$\boldsymbol{Z(G_i^*)} = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}}$$

Where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features, and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \qquad\qquad S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2}$$

The G_i* spatial statistics have been used in a limited number of studies to identify a tendency for positive spatial clustering and distinguish between hotspots and coldspots (Songchitruksa and Zeng, 2010; Aghajani et al., 2017; Sipos, 2017).

Manepalli et al. (2011) stated that the ability of the G_i^* spatial statistic method in identifying hotspots is similar to the KDE method. However, the Gi* spatial statistic can substantially identify the spatial relationships with the concentration of weighted spatial features (Sipos, 2017). Furthermore, the G_i^* spatial statistic identifies statistically significant hotspots and coldspots while the kernel density function calculates the density of features (Aghajani et al., 2017). The G_i^* spatial statistical method is more suitable to identifying spatial patterns of events because it can locate hotspots and coldspots on a global scale and distinguish cluster structures of high or low concentration among local observations (Songchitruksa and Zeng, 2010). Therefore, in this research, the analysis of clustering will be carried out using the G_i^* spatial statistics after projecting crash data in ArcGIS.

SUMMARY

Several significant considerations were found through the review of the literature. Most of these relate to associated factors preceding crashes involving teen drivers, and some of these relate to spatial analysis techniques used to analyze crashes involving teen drivers.

- According to Curry et al. (2011), the critical reason for almost all teen-involved crashes was driver error, and in more than three-quarters of these crashes teen drivers committed the error. Singh (2018) categorized these errors into recognition errors, decision errors, performance errors, and non-performance errors. This consideration represents an important sign for conducting more research on associated factors and effective countermeasure to improve teen driving safety.
- Peek-Asa et al. (2010) indicated that teen-involved crashes on rural roads were five times more likely to cause fatal or severe injuries than those that occurred on urban roads. Most of those crashes that occurred on rural roads were single-vehicle crashes, occurred late at night, involved a failure to yield the right-of-way, or because of crossing the centerline of undivided roads.
- Some studies revealed that the distinguishing characteristics of teen drivers associated with the high-rate of crashes are inexperience, which has negative impacts on hazard

perception, high risk-taking behaviors such as driving at night, when they are fatigued and drowsy, speeding, under the influence of drugs or alcohol, or distracted while driving (Shinar, 2017; OECD, 2006; Carney et al., 2016; Huestegge et al., 2010; Crundall et al., 2012).

- Hellinga et al. (2007) described the vehicles that were driven by teenagers and stated that even though most of the parents realized the importance of major safety criteria in the vehicle, they mostly chose vehicles for their teens based on economic features, not safety features. That means cars driven by teens tend to be smaller, older and provide less advanced safety technologies.
- Several studies concluded that compared to experienced drivers, novice drivers had higher hazard response times, greater speed and steering variability, and higher crash rates in inclement weather, at horizontal curves, and at intersections (Mueller and Trick, 2012; Li et al., 2015; Scialfa et al., 2011; Borowsky et al., 2010; Borowsky et al., 2009). Therefore, determination of locations and/or times that impose critical situations on novice drivers need spatial analysis for crashes involved this group of drivers.
- Ouimet et al. (2015) reviewed several studies on fatal crashes and nonfatal crashes involving teen drivers that compared the performance of teen drivers with passengers present and without passengers. Studies on fatal crashes indicated an increased risk for teen drivers with at least one passenger compared with solo driving. This increased risk became higher for teen drivers with two or more passengers.
- Baker et al. (2006) and Williams et al. (2016) revealed that the GDL program is associated with reductions in teen driver crashes. Baker et al. (2006) found that the most effective GDL programs were programs that included age restrictions, with at least three

months waiting time before intermediate stage policy, nighttime restrictions, and passenger restrictions.

- Several methods have been developed for crash pattern analysis, such as the network screening, the Kernel Density Estimation (KDE) and the Getis-Ord (Gi*) spatial statistics. Aghajani et al. (2017) stated that the Gi* spatial statistics technique is preferred because it can identify statistically significant traffic crash hotspots and coldspots.

The considerations found from the literature reported herein represented useful tools in developing this research through designing the study, presented in Chapter III and selecting countermeasures and analysis techniques, presented in Chapter IV.

CHAPTER III. RESEARCH APPROACH

To accomplish the objectives of the study, a work plan including two paradigms were followed, as shown in Figure 1. In the first paradigm (Phase 1), the descriptive analysis were carried out from the perspective of three major factors: crash, driver, and vehicle while in the latter paradigm (Phase 2), the evaluation of crashes was conducted to define the presence of abnormal clusters of crash patterns and model the impacts of factors associated with the crashes using spatial statistic techniques in ArcGIS.

DESCRIPTIVE ANALYSIS

In this phase, a dataset of seven years (2010-2016) of traffic crashes involving teen-drivers in Kansas was used. Different characteristics associated with crashes and the correlation between characteristics were analyzed. These characteristics are listed under three major related factors shown in Figure 6 and illustrated below:

- Crash characteristics – which included:
    - Type of crashes in terms of number of vehicles involved, the presence of (pedestrian, cyclists, animals, and fixed objects), and type of impacts (rollover, rear-end, head on, side-impact, and sideswipe);
    - Crash conditions with regard to the severity of the crash, weather conditions, light conditions, and time and date of crashes;
    - Crash locations not only in respect of counties and KDOT districts, but also in respect of roadway features such as intersections and horizontal curves, and roadway locations such as urban and rural;

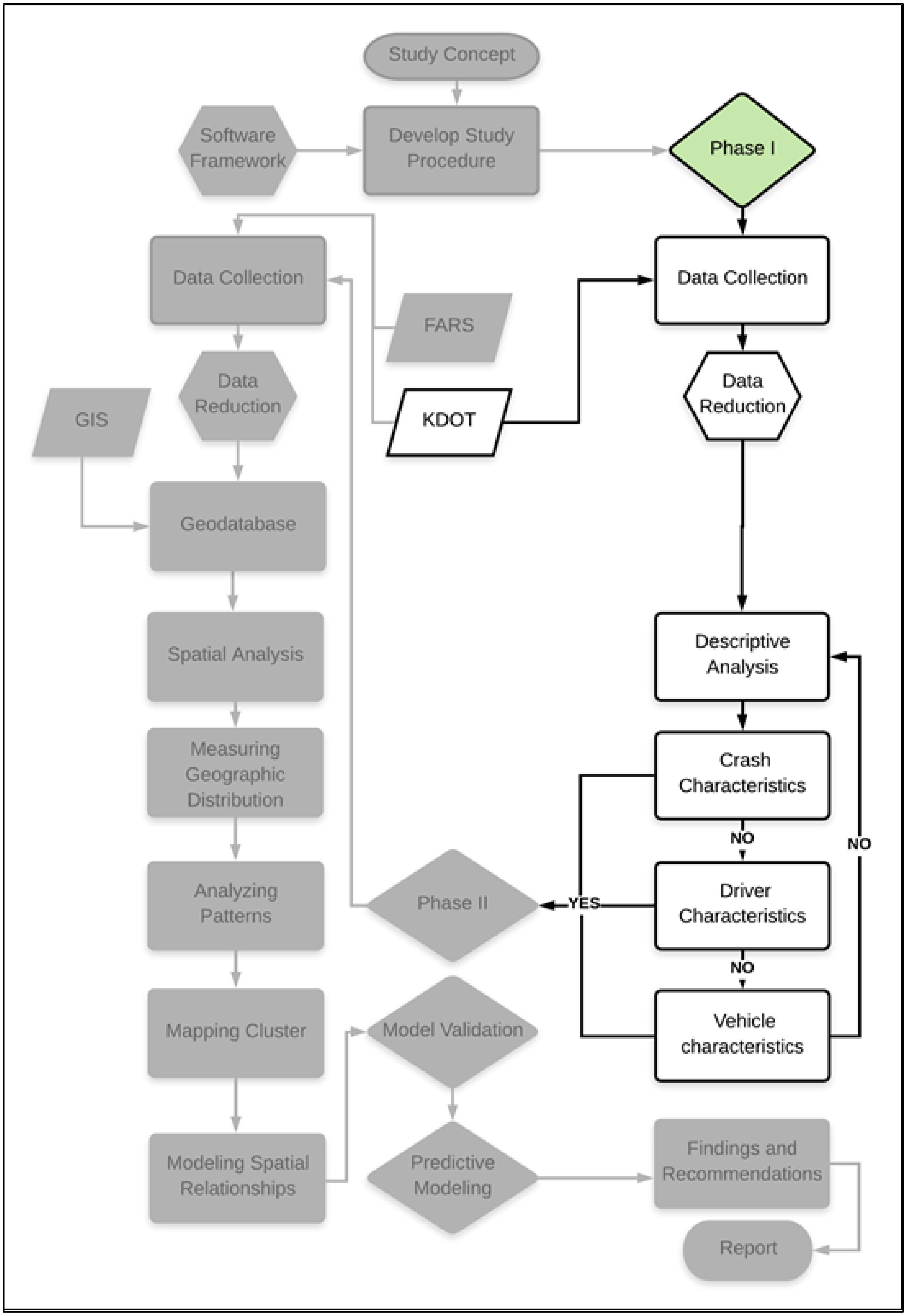


Figure 6. Descriptive Analysis Phase

• Driver characteristics – considers driver characteristics such as age, gender, and risky behaviors such as DUI, safety equipment usage such as seatbelt and airbag, and involving in distracting activities.

• Vehicle characteristics – focuses on the types and model year of vehicles involved in the crashes.

The outcomes of the descriptive analysis revealed crash frequencies by each of the factors listed above. The highest number of crashes in each factor was represented as a safety concern and was considered in the spatial analysis phase.

SPATIAL ANALYSIS

The spatial analysis represents the engine that drives research applications of GIS (Longley, 2005). Spatial analysis is a workflow that provides an approach to solve a problem, which contains specific steps starts with asking questions and closes with making decisions, as shown in Figure 7 (Esri, 2018a).

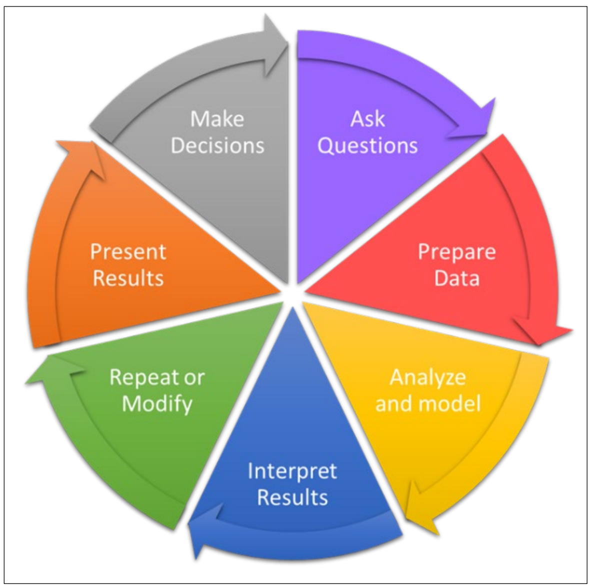


Figure 7. The Spatial Analysis Workflow, Adapted from (Esri, 2018a)

To find out why things are where they are and how things are related when dealing with spatial data, using spatial statistic functions that are added to ArcGIS for analysis is the most appropriate tool. Wade and Sommer (2006) define spatial statistics as a field of study concerning statistical methods that utilize space and spatial relationships directly in their mathematical computations. Spatial statistics could be distinguished from other statistical methods by its concern with features which are locate close to each other in space and tend to share similar characteristics (Longley, 2005). The spatial statistics in ArcGIS are a set of techniques used for describing and modeling spatial distributions, facial patterns, processes, and relationships (Esri Events, 2017b). Mapping crashes by spatial statistics is a visual statistic that can show where clustering is and where actions are needed to meet safety criteria. Mapping hotspots and coldspots of traffic crashes are useful to find locations that meet safety criteria and to visualize the relationship between associated factors in those areas (Mitchel, 1999).

In this phase, the geospatial data were analyzed by spatial statistic functions in ArcMap to identify patterns and map clusters and model spatial relationships for teen driver-related crashes according to the steps shown in phase II in Figure 1. Based on the spatial parameters, besides analyzing the entire state, a smaller location on the unified school district level (e.g., USD 259) was selected as an example to conduct an in-depth investigation to show the utility of the methodology for different sized areas. The data were used to: assess the locations associated with hotspots, identify parameters associated with hotspots, conduct statistical analysis to determine significant factors affecting the rate of traffic crashes in the hotspots, and provide recommendations for selecting appropriate countermeasures that could have crash reduction potential is applied systemically. The applicable statistical functions in spatial statistics toolbox in ArcGIS are grouped into four toolsets that briefly described herein (Fischer and Getis, 2009):

Measuring Geographic Distribution

The tools in this toolset are basic descriptive statistics that are used as a starting point in the spatial analysis process (Pimpler, 2017) to help summarize the main characteristics of a spatial distribution through tools that could be viable in this study, such as: Directional Distribution (or Standard Deviational Ellipse), and Mean Center.

Directional Distribution (or Standard Deviational Ellipse)

This tool measures how geographic features (features are crashes in this study) were distributed spatially around their geometric center (mean), and how was the dispersion and orientation of features over time (Fischer and Getis, 2009). Mapping this trend for a set of features might identify a relationship to specific physical features such as educational centers or location of bars. The axes of the ellipse can be measured by calculating the standard distance of features in the east-west (x-axis) direction and north-south (y-axis) direction separately. This ellipse is commonly called the Standard Deviational Ellipse because the axes of the ellipse are defined by calculating the standard deviation of the x-coordinates and y-coordinates of feature locations from the mean center (Esri, 2014b), as shown in the equation (1) and (2) below:

$$SD_x = \sqrt{\frac{\sum_i (x_i - \bar{X})^2}{n}} \tag{1}$$

$$SD_y = \sqrt{\frac{\sum_i (y_i - \bar{Y})^2}{n}} \tag{2}$$

Where: SD_x and SD_y are the standard distances for the x- and y-axes, x_i, and y_i are the coordinates for feature i, $\bar{X}$ and $\bar{Y}$ are the Mean Center for the features, and n is the total number of features as shown in Figure 8.

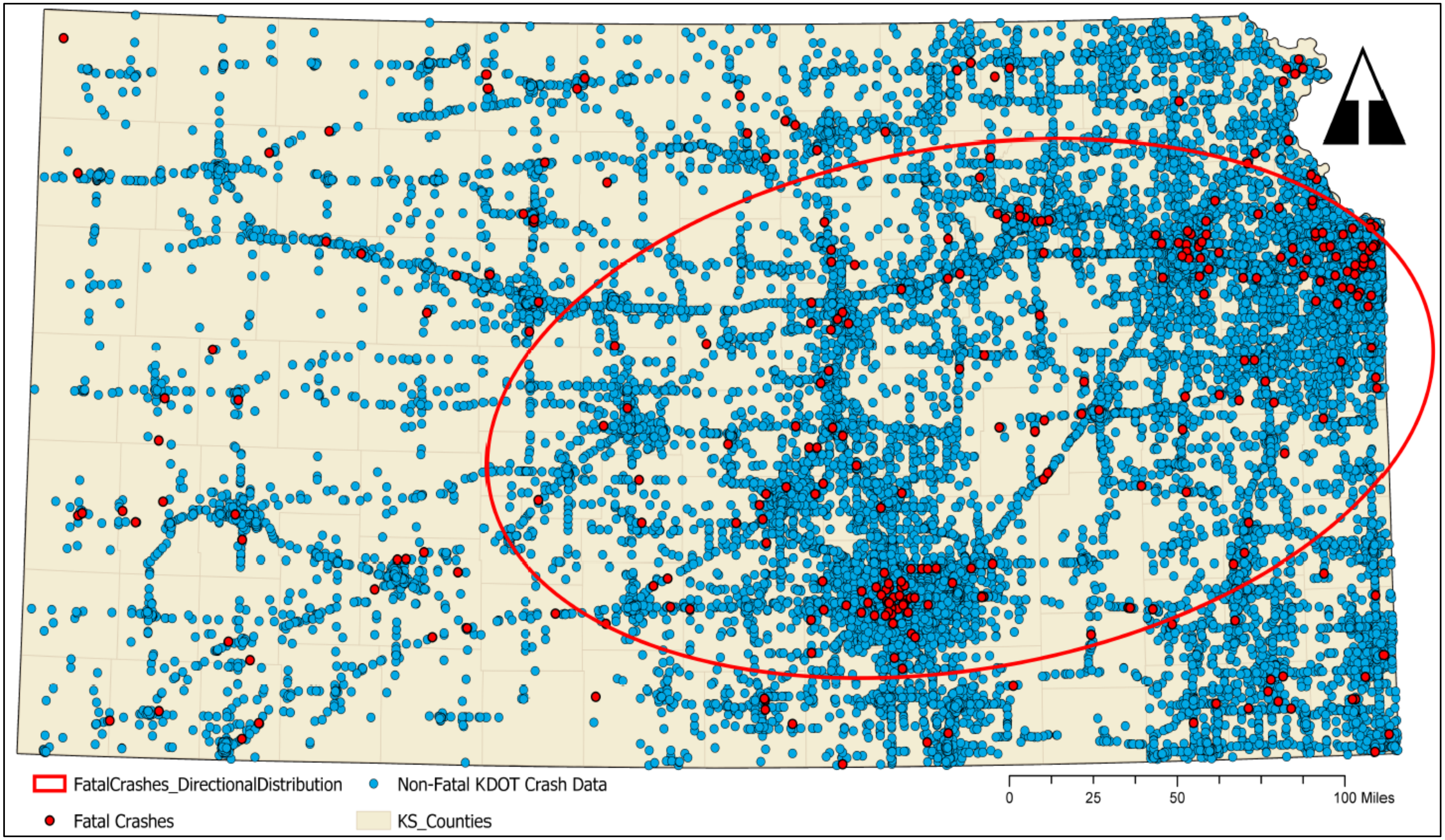


Figure 8. Example of the Directional Distribution

To determine the orientation of the ellipse, the rotation angle of the ellipse (θ) calculates from 0º (north for the y-axis) and is given by equation (3):

$$\tan\theta = \frac{\left(\sum_{i=1}^{n} \acute{x}_i^2 - \sum_{i=1}^{n} \acute{y}_i^2\right) + \sqrt{\left(\sum_{i=1}^{n} \acute{x}_i^2 - \sum_{i=1}^{n} \acute{y}_i^2\right)^2 + 4\left(\sum_{i=1}^{n} \acute{x}_i \, \acute{y}_i\right)^2}}{2\sum_{i=1}^{n} \acute{x}_i \, \acute{y}_i} \tag{3}$$

Where $\acute{x}_i$ and $\acute{y}_i$ are the deviations of the *xy*-coordinates from the Mean Center.

Mean Center

The Mean Center is a measure of central tendency that identifies the geographic center for all features (e.g., crashes) in the study area. This tool is convenient for tracking or comparing changes in the distribution of type of crashes involving teen drivers. It is used to answer questions like "Where are the crashes involving teen drivers centered at daytime and nighttime and how did it move over time?" The mean center in the study area is the average of x–coordinate value and y-coordinate value of all the features, which can be calculated by equation (4) and (5) shown below (Esri, 2014d):

$$\bar{X} = \frac{\sum_i^n x_i}{n} \tag{4}$$

$$\bar{Y} = \frac{\sum_i^n y_i}{n} \tag{5}$$

Where: x_i and y_i are the coordinates for feature i, $\bar{X}$ and $\bar{Y}$ are the Mean Center coordinate values, and *n* is the total number of features.

Analyzing Patterns

The Analyzing Patterns toolset comprises different tools that assist in understanding broad spatial patterns and trends of crashes involving teen drivers. The tools in this toolset are inferential statistics that use statistics to measure clustered features and facilitate comparing crash patterns for different crash types or tracking the changes in their trends over time displayed on maps (Esri, 2014f). Using statistics to measure patterns means comparing observed distribution to a hypothetical random distribution of the same number of observations over the same area statistically (Mitchel, 2005). The tools in this toolset help to evaluate whether features or their corresponding values form a clustered, dispersed, or random spatial pattern (Pimpler, 2017). The most common tools in this toolset are the average nearest neighbor, high/low clustering (G_i*), spatial autocorrelation (global Moran's I), and multi-distance spatial cluster analysis (Ripley's K function).

Average Nearest Neighbor (ANN)

This calculates the average distance between each feature and its nearest neighbor to determine the difference or ratio between the average (mean) range of observation and the expected average (mean) distance for the hypothetical random distribution (Mitchel, 2005). The average distance for the observed distribution ($\bar{d}_o$) and the expected average distance for a random distribution ($\bar{d}_e$) of features can be calculated by equation (6) and (7), respectively:

$$\bar{d}_o = \frac{\sum_i c_i}{n} \tag{6}$$

$$\bar{d}_e = \frac{1}{2\sqrt{n/A}} \tag{7}$$

Where: c_i is the distance between feature I and its nearest neighbor, A is the area, and n is the total number of features.

When the observed average distance and the expected average distance are equal, the difference between them is zero, which means the observed distribution is random. When the difference is less than zero, the pattern exhibits clustering, but when the difference is greater than zero, the trend is dispersion. However, when the ratio between the average distance of an observation and the expected average distance (which called Nearest Neighbor Ratio) was calculated, the resultant threshold value becomes one instead of zero. In such a case, when the ration is less than one, the data are clustered and if the ration is greater than one, the data are dispersed. This tool, in this study, will be used to compare the distribution of two different crash types or crashes for two different driver characteristics to find out which one is more clustered than the other.

High/Low Clustering (Getis-Ord General G)

The Global G measures the concentrations of high or low values of features for a study area by comparing high or low values measured within a specified distance to those values over the entire study area (Mitchel, 2005). This tool can be used to compare the pattern of different types of crashes in a small scale like a city or in a larger size like a county, a district, or a state to find out whether areas with high crash numbers or low crash numbers are clustered or dispersed. This tool uses neighbor features based on a specified distance. If the neighbor feature is within the specified distance of the target feature, it is assigned a weight of 1; otherwise it is assigned a weight of 0. The observed General G can be calculated by equation (8) below:

$$G_o\,(d) = \frac{\sum_i \sum_j w_{ij}\,\left(x_i.x_j\right)}{\sum_i \sum_j\,\left(x_i.x_j\right)}, \forall j \neq i \qquad (8)$$

Where x_i and x_j are attribute values for features i and j, w_{ij} is the spatial weight between feature i and j, and n is the total number of features, and $\forall\, j \neq i$ indicates that feature i and j cannot be the same feature.

To interpret what the G value means, it needs to be compared with the expected G value for a random distribution at the specified distance, which can be calculated by equation (9):

$$G_E\ (d) = \frac{\sum_i \sum_j w_{ij}}{n(n-1)}, \forall j \neq i \tag{9}$$

To test whether the observed G value is significantly different than a random distribution at a specified confidence level, the calculated z–scores provides the evidence. The z–score is calculated by equation (10):

$$Z_G\ (d) = \frac{G(d)_o - G(d)_E}{SD_{G(d)}} \tag{10}$$

Where: $SD_{G(d)}$ is the standard deviation for the expected G.

If the z-score value is positive, the observed General G value is larger than the expected General G value, indicating that clustering of high values for the attribute is occurring. If the z-score value is negative, the observed General G value is smaller than the expected value, which indicates that low values are clustered in the study area. The General G tool is recommended when values (number of crashes in this research) were distributed relatively evenly across the study area, requiring a test to determine if any spikes of high or low values are statistically different than their neighbors (Esri, 2018a).

The study area (Kansas) is divided into different classified polygons, and these polygons could be districts (six KDOT districts), counties (105 counties), USDs (286 USDs), or census tracks. The input feature class has to have a value, which could include the number of fatal or non-fatal crashes in each county or USDs. Therefore, before applying this tool, the number of fatal and non-fatal crashes in each feature needs to be aggregated. The spatial join tool (see Figure 9 for example) is the best option for this purpose. Herein, the output of this spatial join is a polygon, and each polygon represents a county.

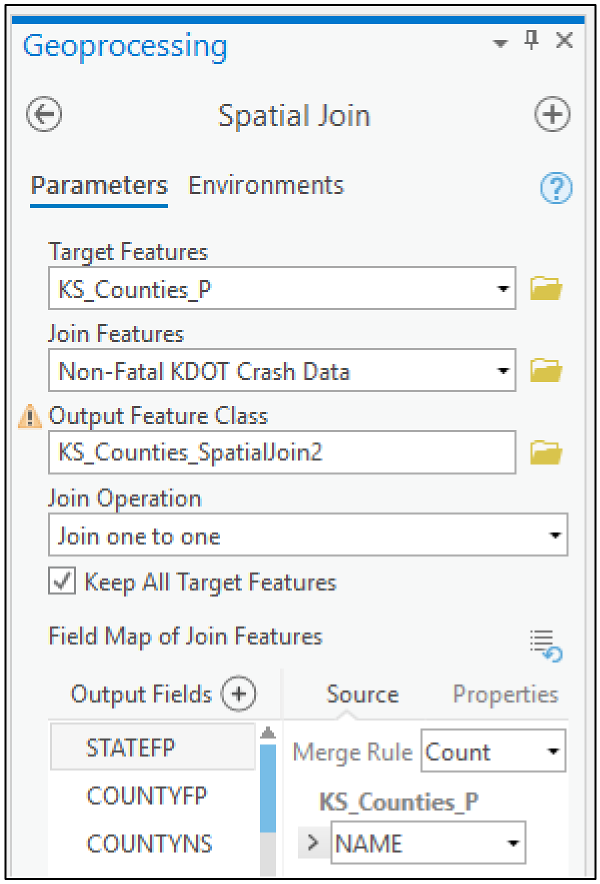


Figure 9. Spatial Join Tool Application

The resultant field (Join_Count) from the spatial join tool containing the number of fatal and non-fatal crashes in each county becomes the Input Field for the High/Low Clustering (Getis-Ord General G) analysis separately, as shown in Figure 10. This method commonly referred to as the Join Count Statistic, which is used with areas' which included nominal data (Mitchel, 2005), such as traffic crashes classified into fatal and non-fatal types, in our case. In the High/Low Clustering (Getis-Ord General G) stage, the software asks for the conceptualization of spatial relationships from a list of provided options. The conceptualization field generally defines the qualified neighborhoods around any individual county. Most of the alternatives identify the relationships in terms of distance and this type of conceptualizations are appropriate when the analysis is based on features like points.

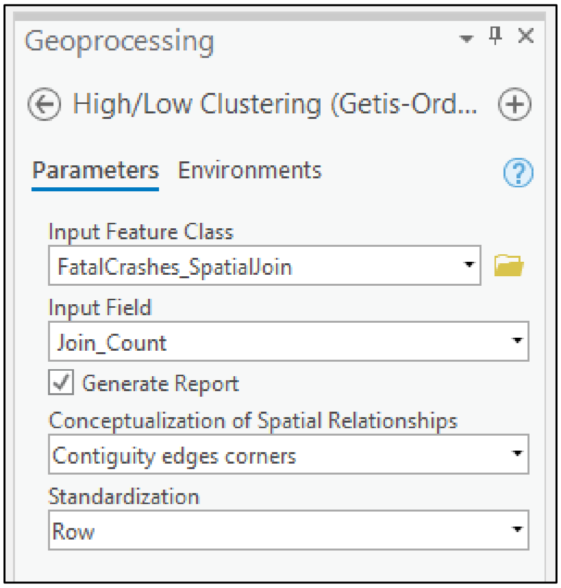


Figure 10. The High/Low Clustering Application

Since this analysis was conducted based on polygons (counties) with different sizes and shapes, the distance relationship was eliminated, and the contiguity (Contiguity_Edges_Corners) option was selected. This means counties that touch one another were qualified and counted as neighbors. When selecting this option, the distance threshold becomes irrelevant.

For the standardization field, there are two options, none and row. Whenever the polygon contiguity conceptualization was selected, row standardization for tools that have the row standardization parameter was selected. Because the row standardization accounts for the fact that some counties have more neighbors than others and there was a desire not to influence the relative impact of the statistical analyses. Otherwise, the number of neighbors was likely to be a function of the size of the counties, which means smaller counties counted for more neighbors and also counties located at the edge of the state count for fewer neighbors.

The General G can tell whether clustered or dispersed exist in the study area, but to know whether similar values are clustered or dispersed, the Moran's I is the proper tool to apply.

Spatial Autocorrelation (Global Moran's I)

Global Moran's I measures whether similar values are clustered or dispersed by calculating the differences between the target feature and the mean on the one hand, and between each feature and the mean on the other hand, and then both differences will be multiplied in turn to get a "cross-product" value. A high or low cross-product value indicates that nearby features have similar or dissimilar values, respectively. When feature with similar values are located close to each other, the pattern is considered to be clustered. When features with dissimilar values are located close to each other, the pattern is considered to be dispersed. Mathematically, the Global Moran's I values are ranged between -1 and 1. When the value of the Moran's I is positive, that indicates clustering of values. Conversely, when the value of the Moran's I is negative, that indicates the dispersion of values.

To test the significance of the Moran's I value statistically, ArcGIS calculates the z-score using the expected Moran's I value (I_E) for a random distribution and observed Moran's I value (I_o), as shown in equation (11), (12) and (13) below (Mitchel, 2005):

$$I_o = \frac{n \sum_i \sum_j w_{ij}(x_i - \bar{X})(x_j - \bar{X})}{\sum_i \sum_j w_{ij} \ \sum_i (x_i - \bar{X})^2} \tag{11}$$

$$I_E = \frac{-1}{n-1} \tag{12}$$

$$Z_I = \frac{I_o - I_E}{SD_{I_E}} \tag{13}$$

Where: x_i is an attribute for feature i, $\bar{X}$ is the mean of the corresponding attribute, $w_{i,j}$ is the spatial weight between feature i and j, n is the total number of features.

The major point that distinguishes the Global Moran's I from the General G tool is that the Spatial Autocorrelation (Global Moran's I) tool also an adjudicated tool when both the high values and the low values cluster. Conversely, with using The High/Low Clustering (Getis-Ord General G) tool when both the high and low values cluster, they be likely to cancel each other out (Esri, 2014f). When points are analyzed, the Spatial Autocorrelation (Global Moran's I) tool also asks for the threshold distance to run the analysis. The threshold distance represents the size of the area around each point (crash in the case of this research) to constitute a neighborhood so that all features should have at least one neighbor for the analysis to be reliable, and no feature should have all other features as a neighbor. The Spatial Autocorrelation tool uses an algorithm to determine the threshold distance depending on the study area size (Esri, 2014f). This distance is a significant input value in the Spatial Autocorrelation (Global Moran's I), Hot Spot Analysis (Getis-Ord *Gi**), and Cluster and Outlier Analysis (Anselin Local Moran's I) tools.

To check whether the threshold distance value is appropriate, two integrated tools in ArcGIS Software are used for verification. These tools are the Calculate Distance Band from Neighbor Count and the Incremental Spatial Autocorrelation. The Calculate Distance Band from the Neighbor Count tool calculates the minimum, the maximum, and the average distance of a feature to the specified number of the nearest neighbor while the Incremental Spatial Autocorrelation tool runs the Spatial Autocorrelation (Global Moran's I) tool for a series of increasing distances to measure the intensity of spatial clustering for each distance. (Esri, 2014f). Recently, Esri added a new tool called Optimized Hot Spot Analysis, which is a combination of the Incremental Spatial Autocorrelation tool and the Hot Spot Analysis (Getis-Ord *Gi**). The Optimized Hot Spot Analysis tool is able to calculate the threshold distances directly with the Hot Spot Analysis, and this eliminates all the steps mentioned above.

However, in the analysis process, the resultant field (Join_Count) from the spatial join tool implemented in the previous section (General G) containing the number of crashes in each county becomes the Input Field for the Spatial Autocorrelation (Global Moran's I) analysis, as shown in Figure 11. The output of Global Moran's I tool is a Portable Document Format (pdf) report which includes parameters such as z-score, p-value, and Moran's I values of the analyzed dataset.

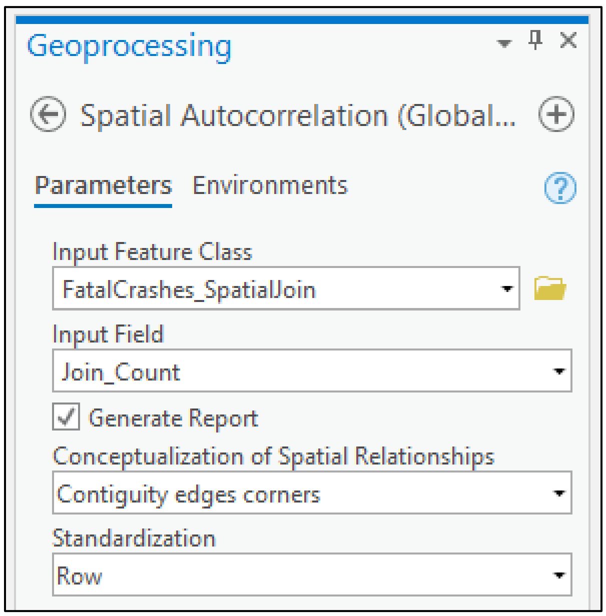


Figure 11. Spatial Autocorrelation Analysis Application

Multi-Distance Spatial Cluster Analysis (Ripley's K Function)

Ripley's K function is another way to analyze the spatial pattern of crash point data. This tool identifies spatial clustering and dispersion for features at a series of distances, which represent threshold distances established around each feature (Fischer and Getis, 2009). The K function helps to display how the clustering or dispersion of crashes changed when the size of the neighborhood changed, as shown in a line graph in Figure 12. The blue diagonal line represents the expected pattern when the feature randomly distributed while the red curved line represents the observed pattern. The clustered and dispersed patterns are located at a distance when the curve line goes above or below the expected pattern, respectively. The clustering and dispersion are statistically significant when they are located outside of the confidence range (higher confidence and lower confidence).

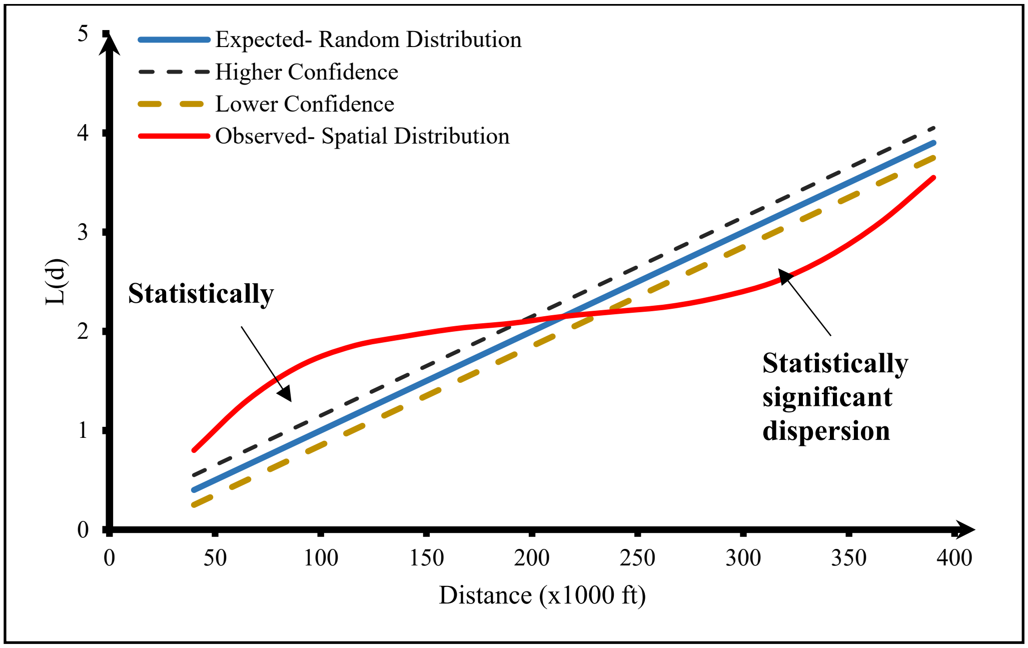


Figure 12. Components of the K Function Output. Adapted from (Fischer and Getis, 2009)

The K Function for any specified distance can be calculated using equation (14) below:

$$(d) = \frac{A}{n^2} \sum_i \sum_j I_{ij} \ d_{ij}, \forall j \neq i \tag{14}$$

Where: *I* is feature's weight, which is either 1 when it locates within the specified distance (*d*) from the targeted feature or 0 when it locates outside the specified distance.

When plotting the resulting values on a chart to display the pattern, as the distance increase on the x-axis, the K Function values on the y-axis get very large and makes the chart virtually unreadable. To improve the chart's view and reduce the height of y-axis, the raw value of K Function could be transformed to $L\ (d)$ value using equation (15) (Mitchel, 2005; Esri, 2014e).

$$L(d) = \sqrt{\frac{A \sum_i \sum_j I_{ij}\ d_{ij}}{\pi n(n-1)}}, \forall j \neq i \tag{15}$$

The K Function could also be used to compare the distribution of different type of crashes or different age of drivers and find out which one is more clustered than the others. Or it could be used to determine whether the distribution of crashes involving teen drivers in the study area and then the distribution of the population.

This tool was not used in this research for three major reasons: the application of this tool was not adding any new results to the analysis, the graphical display of results has been eliminated from the ArcGIS Pro software, and the Incremental Spatial Autocorrelation tool, which runs the Spatial Autocorrelation (Global Moran's I) tool for a series of increasing distances to measure the intensity of spatial clustering for each distance, has been added to a new hot spot analysis tools (Optimized Hot Spot Analysis), which limits the necessity of applying the K Function tool.

Mapping Clusters

In the analysis pattern, the tools try to help answer whether there are statistically significant spatial patterns such as clustering or dispersion in the study area, whereas the tools in the mapping clusters are trying to visualize locations and extent of clusters (hotspots and coldspots) and answer where spatial clusters are located (Fischer and Getis, 2009). Logically, answering these types of questions is essential and represent a milestone to examine the potential contributing factors of hotspots and coldspots, and to identify locations have the priority is in taking improvement actions. Therefore, this toolset is the most commonly used and well-known toolsets among the spatial statistics tools (Pimpler, 2017). The mapping clusters toolset includes two tools that identify locations where spatial clustering and spatial outliers occurs in the study area. The tools are Hot Spot Analysis (Gi∗), and Cluster and Outlier Analysis (Anselin Local Moran's I).

Hot Spot Analysis (G_i*)

G_i* (pronounced "G–i–star") is a spatial statistic tool that identifies clusters of high values (hotspots) and clusters of low values (coldspots) for a set of weighted features within a specified distance using the G_i* statistics (Fischer and Getis, 2009). The Hot Spot Analysis separates the neighborhood of each feature (including the feature itself) from the study area and finds out if the value of this neighborhood is significantly different from the value of the overall study area. The feature is marked as a hotspot when the value of the neighborhood is statistically significantly higher than the value of the study area or is marked as a coldspot when the value of the neighborhood is statistically significantly lower than the value of the study area at different confidence levels, otherwise marks as random (Esri Events, 2017a). The Gi∗ represents a z-score for each feature in the dataset, and it can be calculated using the equation (16) below (Esri, 2014c).

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}} \tag{16}$$

Where: x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is the total number of features, $\bar{X}$ and S are given by equation (17) and (18).

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{17}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \tag{18}$$

Since the Gi* statistic is a z-score, no further calculations are required, and the resultant Gi* identify statistically significant spatial clusters of high values, which indicate hotspots, and low values, which indicate coldspots. The resultant Gi* tells whether or not to reject the null hypothesis, which states complete spatial randomness among the values associated with each feature. The output of this tool includes the p-value as well, which is the probability that a random process formed the observed spatial pattern. When the p-value is smaller than the required level of significance, the null hypothesis could be rejected because the p-value, in this case, indicates a small probability that the observed spatial pattern is the result of random processes (Esri, 2013). The z-scores are standard deviations, and the furthest z-scores from zero are associated with smaller p-values, which are an indication of significant spatial clustering.

Besides the hotspots and coldspots, the Optimized Hot Spot Analysis tool calculates that distance by utilizing its component, the Incremental Spatial Autocorrelation tool. The outputs of the Optimized Hot Spot Analysis tool are z-scores and p-values in different levels of significance, which represent the determination of what is a hotspot and what is a coldspot. High z-scores indicate statistically significant spatial hotspots, and low z-scores indicate coldspots. With p-values smaller than the required level of significance, the null hypothesis, which is complete spatial randomness, could be rejected.

Because of its efficiency and its comprehensiveness in comparison to the typical Hot Spot Analysis tool, the Optimized Hotspot Analysis tool was used to identify hotspots and coldspots in the study area. The window of the Optimized Hotspot Analysis tool runs incident features and do not require any attribute values, as shown in Figure 13. The other two significant parameters in its window are “Incident Data Aggregation Method” and the “Bounding Polygons Defining Where Incidents Are Possible.” The Incident Data Aggregation Method calculated all incident points (crashes in our case) in each specified polygon. The provided options are:

- Count incidents within a fishnet grid, which creates an appropriate fishnet (grid) mesh that counts the number of crashes located in each cell.
- Count incidents within a hexagonal grid, which creates an appropriate hexagonal polygon mesh that counts the number of crashes located in each hexagon. The dimensions of hexagons are displayed by width (length on x-axis), and height (length on y-axis).
- Count incidents within an aggregation of polygons, which requires that the polygons containing crashes is provided. An example of this could be counties containing the number of crashes.
- Snap nearby incidents to create weighted points, which can be used to aggregate nearby crash points, so each point can be weighted by the number of crashes that were involved.

Because the existent polygons inside Kansas (which could be divided to districts, counties, USDs, or census tracks) are not identical in sizes and shapes, the study area had to be divided into identical polygons to avoid size bias. Therefore, the “Count incidents within the hexagon grid” option was selected, which divides the study area into several similar hexagonal polygons. The shape of each hexagon reduces sampling bias caused by the edge effects of the grid shape (Esri, 2014c). Each of the polygons contains the number of crashes located in that polygon under a column (Counts) created by the software. In the “Bounding Polygons Defining Where Incidents Are Possible” field, the shapefile of Kansas boundary was uploaded to identify the border of the study area. The output of the Optimized Hotspot Analysis is a map which shows every statistically significant coldspots and hotspots in the study are.

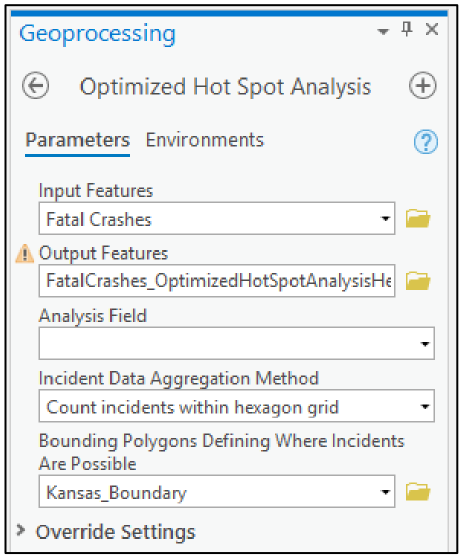


Figure 13. The Optimized Hot Spot Analysis Window

Cluster and Outlier Analysis (Anselin Local Moran's I)

This tool is the local version of Moran's I, which identifies clusters of high or low values as well as spatial outliers for a set of weighted features. This tool compares each value in the pair features (target feature and neighboring feature) to the mean value for all-inclusive features in the study area. In other words, the Cluster and Outlier Analysis tool separates the neighborhood of each feature (excluding the feature itself) from the study area and determines if the value of this neighborhood based on this feature is significantly different from the value of other neighborhoods and is the value of the feature different from other features in the study area. The Local Moran's I (I_i) can be calculated using the equation (19) below:

$$I_i = \frac{(x_i - \bar{X})}{S^2} \sum_j w_{ij}(x_j - \bar{X}) \tag{19}$$

$$S_i^2 = \frac{\sum_j \left(x_j - \bar{X}\right)^2}{n-1} - \bar{X}^2, \forall j \neq i \tag{20}$$

Where: S^2 is the variance, which is given as in equation (20), x_i is an attribute for feature i, $\bar{X}$ is the mean of the corresponding attribute, $w_{i,j}$ is the spatial weight between feature i and j, n is the total number of features.

If the resulted value of Local Moran's I is a high positive value, it indicates that the targeted feature is bordered by similar values, which could be either high values or low values. But If the resulted value of Local Moran's I is a very negative value, it indicates that the targeted feature is bordered by dissimilar values (Mitchel, 2005). To test the statistical significance of each value of Local Moran's I at a specified confidence level, the z-score gives the last decision of rejecting or not rejecting the null hypothesis. The z-score is calculated by dividing the difference between the expected value (I_{i_E}) and observed value (I_{i_o}) of the local Moran's I by the square root of the variance, as shown in the equation (21) below (Mitchel, 2005):

$$Z_{I_i} = \frac{I_{i_o} - I_{i_E}}{\sqrt{var\ (I_i)}} \tag{21}$$

Where: the expected Local Moran's I is given as in equation (22)

$$I_{i_E} = \frac{-\sum_j w_{ij}}{n-1} \tag{22}$$

A high positive value of the z-score indicates that the targeted feature is bordered by features with similar values, which means clusters of either high values or low values (hotspots or coldspots). But the z-score with a negative value indicates that features border the targeted feature with dissimilar values (Esri, 2014a). To put it simply, the outputs of the Cluster and Outlier Analysis categorize into four different types of significant levels (excluding not significant), as shown in Figure 14.

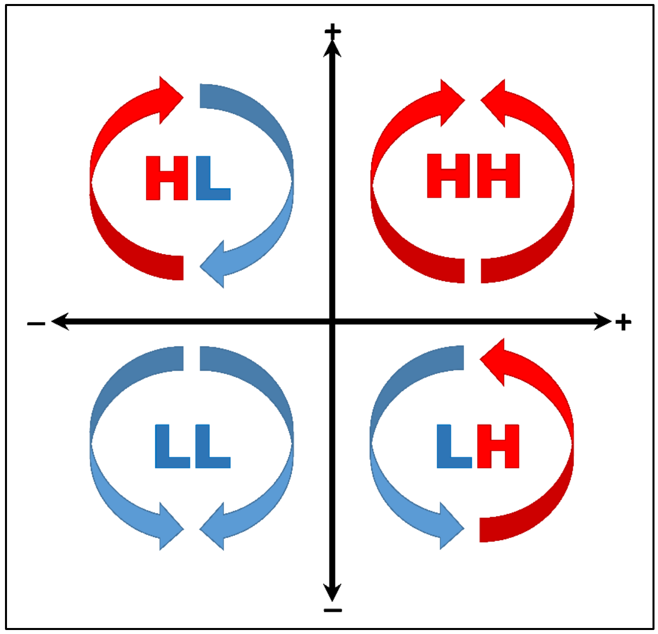


Figure 14. The Quadrants of the Cluster and Outlier Analysis (Local Moran's I)

- High-High (HH) Cluster: Occurs when a feature value is significantly higher than other features, and its neighborhood value is also significantly higher than other neighborhood values, which represent a hotspot.
- Low-Low (LL) Clusters: Occurs when a feature value is significantly lower than other features, and its neighborhood value is also significantly lower than other neighborhood values, which represent a coldspot.
- High-Low (HL) Outlier: Occurs when a feature value is significantly higher than other features but its neighborhood value is significantly lower than other neighborhood values.
- Low-High (LH) Outliers: Occurs when a feature value is significantly lower than other features but its neighborhood value is significantly higher than other neighborhood values.

The Cluster and Outlier Analysis tool does exactly what the Hot Spot Analysis tool does besides determining outliers. So the question might be asked "why do we use the Hot Spot Analysis tool when the Cluster and Outlier Analysis tool is more comprehensive and does much more?" The reasonable answer is running any one of them, or both of them depends on the research question being asked and generally running both of them provides a check tool and also provides a different view to understand the analyzed data. In addition, the existence of outliers (High-Low and Low-High) in a dataset contradict the first law of geography, which states that near features are similar to each other (Tobler, 1970). An adequate explanation of existing outlier features is that the distribution of those features is not random, and more investigation is required to determine the reasons.

Similar to the Hot Spot Analysis Tool, the recently–added version of the Cluster and Outlier Analysis tool is the Optimized Cluster and Outlier Analysis tool, which provides more parameters and also calculates the optimal threshold distance. Since the feature classes were created in the Optimized Hot Spot Analysis section which already assembled crash numbers in hexagon polygons under the "Counts" column, the same feature class can become inputs for the Optimized Cluster and Outlier Analysis tool, as shown in Figure 15. Inputting the calculated threshold distances for fatal and non-fatal crashes from the Optimized Hot Spot Analysis section are optional. Even if left blank, the tool could run the Incremental Spatial Autocorrelation tool to calculate the threshold distance, and its value is the same as what the Optimized Hot Spot Analysis tool calculated.

Besides the fact that the Optimized Hot Spot Analysis tool and the Optimized Cluster and Outlier Analysis tools use different techniques to identify clusters, the Optimized Cluster and Outlier Analysis tool uses a performance adjustment, which includes three levels of permutations

(Quick, Balanced, and Robust). The permutations are used to compare the Local Moran's I of the analyzed dataset to a set of randomly generated values. This function makes the resultant clusters obtained by the Optimized Cluster and Outlier Analysis tool different and more reliable.

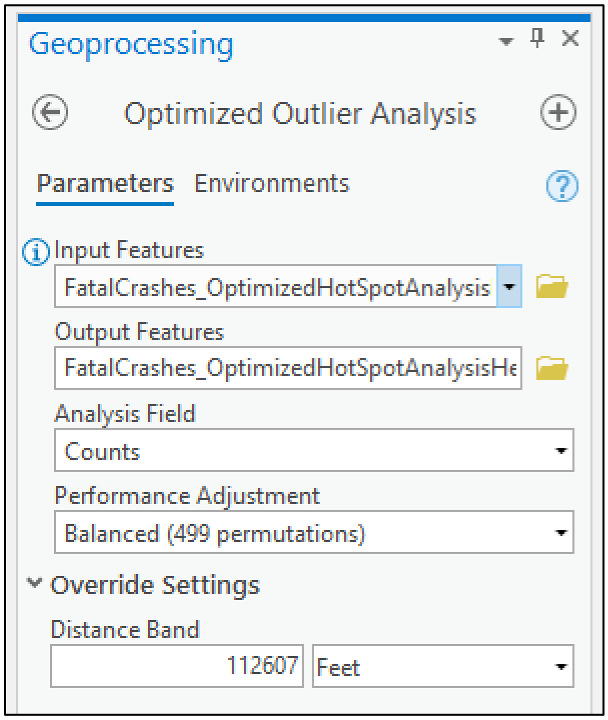


Figure 15. Cluster and Outlier Analysis Application

Beyond analyzing the distribution of geographic features, ArcGIS can be used to identify and measure the relationships between features, which helps to questions like “why do crashes involving teen drivers occur where they do?” or “ where are these crashes more likely to occur in the future?” Analyzing the relationships between factors or variables combined to a spatial feature, such as traffic crashes involving teen drivers is mostly referred to as a modeling spatial relationships (Mitchel, 2005).

Modeling Spatial Relationships

The ArcGIS software provides a toolset for modeling spatial relationships. The most important functions that this toolset provide are modeling, examining, and exploring spatial relationships among features using regression analysis to determine the contributed factors behind observed spatial patterns or to predict spatial outcomes (Fischer and Getis, 2009). The regression analysis has a different type of application. In this study, it can be used to demonstrate the strength of relationships and impacts of several contributed factors that potentially promote negative or positive changes in the number of crashes involving teen drivers. This can help in better understanding these crashes and predict their numbers to make better decisions to improve teen drivers' safety.

Regression analysis attempt to answer most of the *why* and/or *what* questions such as: why are the expected rate of traffic crashes involving teen drivers exceptionally high in particular locations in Kansas? Or what are the potential factors that make some areas have more than the expected rate of traffic crashes involving teen drivers? (Scott and Pratt, 2009). In nonspatial statistical methods that measure relationships, two major assumptions about analyzed data have been confirmed to validate the results, which are independence and randomness among observations in the study area. However, spatial statistical methods do not always hold to these assumptions; events often assumed to be spatially autocorrelated and spatially heterogeneous and nonstationary across the study area (Getis et al., 2005). The salient tools in this toolset are Ordinary Least Squares and Geographically Weighted Regression.

Ordinary Least Squares (OLS)

Shults et al. (2015) define the OLS as a global model that creates a single equation that best describes the data relationships between a response variable and each one of explanatory variables in the study area. The output of the OLS is a single equation that best describes the data relationships between a response variable and each one of the explanatory variables in the study area. Several related variables were prepared for modeling. The selected variables depended on their relativity to the study topic, the availability, and accessibility to the targeted variables. The scope of this topic made obtaining the desired variables a challenging task. The potential eighteen explanatory variables that were believed might affect the dependent variable LNLOCRASH: the number of traffic crashes involving teen drivers from 2010 to 2016 in each county in Kansas are listed alphabetically below:

- Average DVMT on all types of roads (DVMT_ALL);
- Average DVMT on rural non-state roads (DVMT_NONSTATE);
- Average household income (AV_HOUSEH_INCOM);
- Average precipitation in inches (AVG_PRECIPT);
- Miles of all types of roads (ALL_ROAD);
- Miles of rural non-state roads in a county (NONSTATE_RD);
- Number of families whose income is below the poverty level (UNDER_POV_LEV);
- Number of High Schools (HIGHSCHOOL);
- Number of postsecondary schools, e.g., colleges, universities, and other educational centers (POST_SECNDRY);
- Number of workers that have 16 years and over and commuting to work (LNCOMMUT_WORK);

- The average number of non-commercial trucks (AVG_TRUCK);
- The average number of passenger cars (LNLOAVG_PC);
- The population of 16 years and over, who are in the labor force (NLABOR_OVER15);
- The population of 18 to 24 years old that have less than high school graduate degrees (POP18_24NO_HIGHSCH);
- The population of counties (LNPOP);
- The population of Females 16 years and over, who are in the labor force (NFEMALE_OVER15);
- The population of Males 16 years and over, who are in the labor force (NMALE_OVER15); and
- The population of teens (LNLOT_POP).

Since the OLS and GWR are both linear regression methods, the relationship between all of the explanatory variables and the dependent variable needs to be linear; otherwise, the resultant model will perform poorly. A scatter plot matrix graph was used to clarify the relationships among the proposed variables. The variables that had nonlinear relationships or curvilinearity relationships were treated by transforming their values using square roots or logarithm transformations such as the Common Logarithm (log: a logarithm with base 10) and/or Natural Logarithm (ln: a logarithm with base e). For instance, the dependent variable (LNLOCRASH) and the exploratory variable (LNLOAVG_PC) were transformed by applying ln and log to their values while the exploratory variable (NONSTATE_RD) was transformed using the square root and ln.

To identify the exploratory variables that are significant to explain the dependent variable, ArcGIS provides the Exploratory Regression tool. The Exploratory Regression tool is a data-mining tool that assesses every possible combination of the potential explanatory variables entered for OLS models that best explain the dependent variable, as shown in Figure 16.

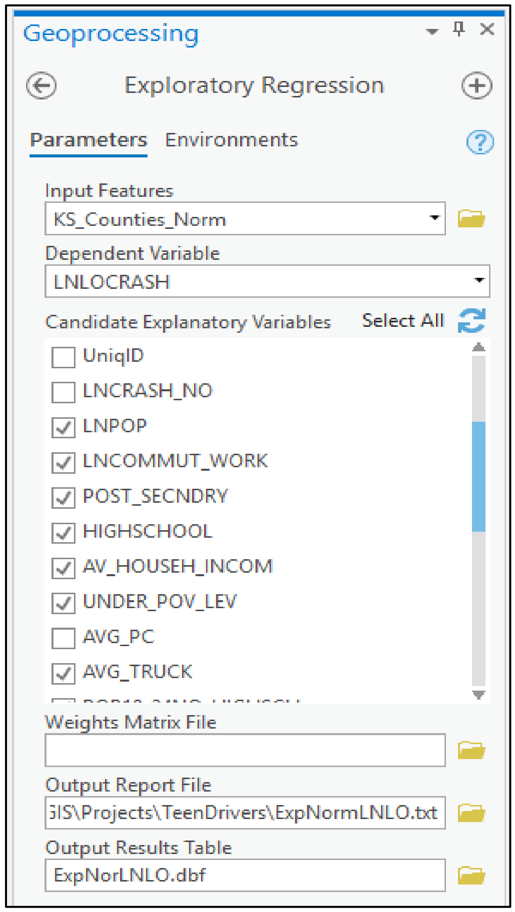


Figure 16. The Exploratory Regression's Application Window

This tool uses OLS and Global Moran's I for its analyses. Its output is a text file not only summarizing significant variables for the OLS model, but also details about any OLS models found that passed all threshold criteria shown below:

- Acceptable Adjusted R Squared;
- p-value Cutoff;

- Variance Inflation Factor (VIF) Value Cutoff;
- Koenker (BP) Statistic p-value;
- Jarque-Bera p-value;
- Joint F-Statistic p-value; and
- Joint Wald Statistic p-value.

The Koenker (BP) Statistic examines whether the explanatory variables have a consistent relationship to the dependent variable in geographic space based on a null hypothesis that the model is stationary and one model can be used throughout the study area. When the Koenker (BP) Statistic is statistically significant, the robust probability is the decisive parameter to determine whether explanatory variables are statistically significant. The VIF measured redundancy among the explanatory variables and their values were less than 7.5, which means the variables were inconsistent in predicting the number of crashes. The Joint F-Statistic and Joint Wald Statistic measured whether the overall model was statistically significant. The Jarque-Bera Statistic tests a null hypothesis that the residuals are normally distributed.

The final OLS model could be expressed in the generic form of an equation as shown below:

$$E\ (y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n + \varepsilon$$

Where *E (y)* is the dependent variable, β is the coefficient, x are the exploratory variables, and ε is residuals.

The OLS regression equation models these relationships precisely when they were consistent across the study area, but when these relationships are heterogeneous across the study area, the regression equation creates an average of the mixed relationships present (Scott and

Pratt, 2009). One of the conventional methods to deal with the regional variation is to incorporate it into the Geographically Weighted Regression (GWR) regression model.

Geographically Weighted Regression (GWR)

The GWR is a local model that creates an equation for every feature in the dataset and the coefficients in the model rather than being global estimates specific to a targeted location (Fischer and Getis, 2009; Brunsdon et al., 1996). The GWR models the relationships that change over the study area by creating a separate equation for each feature using the same explanatory variables applied in the OLS model (Pimpler, 2017).

The GWR is treated in this research as a spatial disaggregation of the OLS. For instance, analyzing the teen-related crashes across Kansas creates a global model, but analyzing those crashes in a KDOT district or a county produce a local model for this area. The OLS regularly produces a single value for the whole study area while the GWR provides different values for different locations in the study area (Fotheringham et al., 2002).

The GWR was applied in a similar manner to the OLS. As shown in Figure 17, the same dependent variable and explanatory variables were entered. However, in the GWR process, there is a model type field that contains three options (Esri, 2018a):

- Continuous (Gaussian), for dependent variables that take a wide range of values;
- Binary (Logistic), for dependent variables that take one of two possible values; and
- Count (Poisson), for dependent variables that take discrete and represents the number of an event.

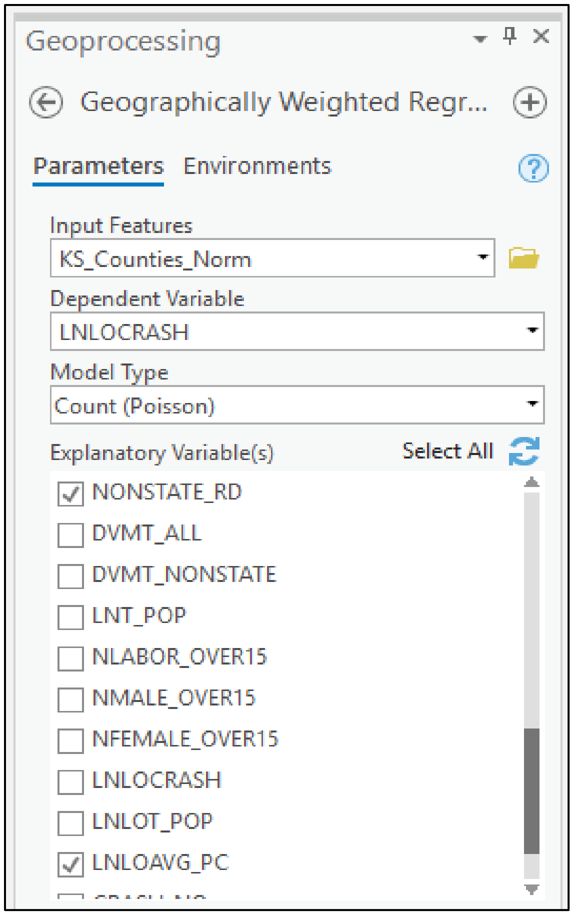


Figure 17. The GWR Application Window

These options are based on how the dependent variable was measured in the dataset. The last option was selected because it is mainly for discrete dependent variables that represent the number of events such as a count of traffic crashes.

Spatial analysis by spatial statistic functions mentioned above can be conducted for different size areas such as countries, states, districts, counties, cities, or zones following the logical steps shown in Figure 18. The remarkableness of implementing spatial statistics for different area sizes is using an appropriate geographic coordinate system and map projection, which is a mathematical method used to portray the curved surface of the earth on a flat surface.

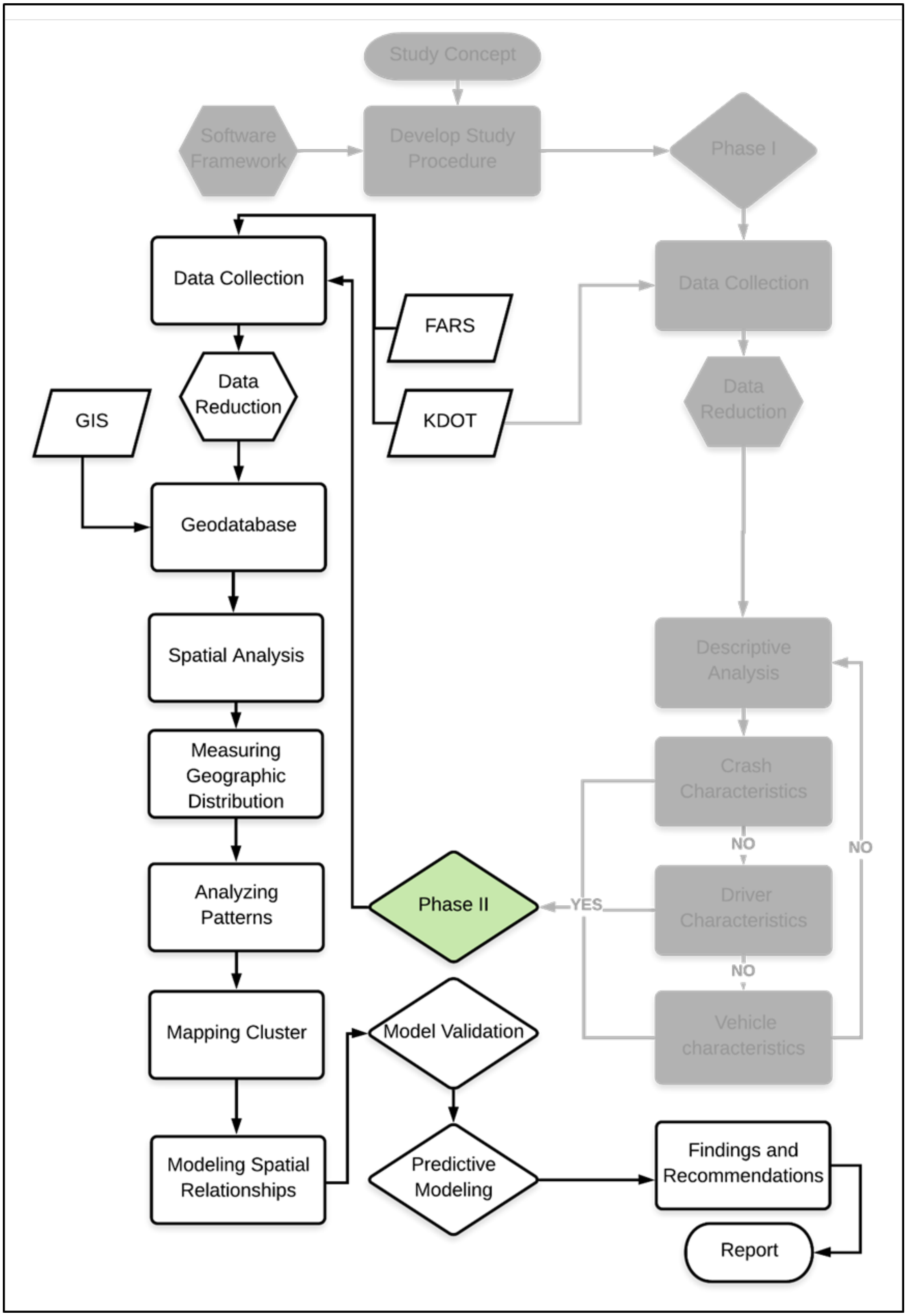


Figure 18. Spatial Analysis Phase

The most popular geographic coordinate systems are, globally, the World Geodetic Survey for 1984 (WGS84) and, nationally, the North American Datums of 1983 (NAD83). The most popular projected coordinate systems used in the state include the NAD83 Kansas Lambert Conformal Conic (LCC) and the Universal Transverse Mercator (UTM). The State of Kansas is located in the UTM Zone 14 North and the UTM Zone 15 North. Therefore, in this research, the NAD83 Kansas LCC coordinate system was used for spatial analysis at the state level. However, the NAD_1983_UTM_Zone_14N or the NAD_1983_UTM_Zone_15N was used for districts and counties level, based on their location.

CHAPTER IV. STUDY AREA AND DATA

STUDY AREA

The state of Kansas is a landlocked Midwestern state located almost in the center of the US, as shown in Figure 19. Kansas is the 15th largest state by area, which has 82,278 square miles of land, the 34th most populated state with a population of 2,911,641 (Kansas.gov, 2018). In Kansas, There are two major metropolitan areas, the Kansas City metropolitan area and Wichita area while the rest of the areas of the state are small cities or rural areas. Kansas is divided into 105 counties. Johnson County represents the most populated county, and Wichita represent the largest city in the state, as shown in Figure 20 and Table 2.

KDOT reported in 2008 that the collective length of all Kansas roads was 140,378 miles (KDOT, 2008). The primary interstate highways in Kansas are I-70 and I-35, as shown in Figure 21. I-70 is a major east-west Interstate Highway that begins at its junction with I-15 at Cove Fort, Utah, and continues 2,151 miles to end in a junction with I-695 near Baltimore, Maryland. About 424 miles of I-70 is located in Kansas and passes through several cities such as Kansas City, Lawrence, Topeka, Junction City, Abilene, and Salina (FHWA, 2018). I-35 is a major cross-country, south-north Interstate Highway that runs from the Mexican border near Laredo, Texas in the south, and continues 1,568 miles to Duluth, Minnesota in the north. About 235 miles of I-35 is located in Kansas, which goes from the Oklahoma border to Kansas City at the Missouri border and passes through several cities such as Wichita, Emporia, Ottawa, and Kansas City (FHWA, 2018).

KDOT is divided into six districts, as shown in Figure 22, and has developed a route classification system based on daily traffic, route continuity, access to major cities, trip length and route spacing, in an effort to better manage and address the diversity of the Kansas state highway system. The system classification divides the state highways into five classes (KDOT, 2008).

- **CLASS A** includes routes that have full access control, permits high-speed travel, and high truck volumes. Examples for this class include the Interstate Highways such as I-70 and I-35. Even though this class represents only eight percent of the state's highway system, they carry more than 40 percent of the traffic volume (21,700 vehicles per day on average).
- **CLASS B** includes non-interstate routes with limited access, high-speed travel, and long-distance truck traffic that serves as the most critical statewide and interstate corridors for travel. Examples for this class include US-50, US-36, and US-400, which carry 5,100 vehicles per day on average. A significant number of out-of-state vehicles use Class B routes.
- **CLASS C** includes routes used usually for regional travel and connects to higher-speed, limited-access roadways. The routes in this class are closely integrated with Class A and B routes in service to all parts of the State. An example for this class is US-77, which it carries 3,800 vehicles per day on average.
- **CLASS D** includes routes usually provide inter-county movement and connect to higher-speed roads. Examples for this class include US-50B, K-16, and K-25, which carry 1,800 vehicles daily on average.

Figure 19. Location of Kansas in the US (Wallpaperama, 2006)

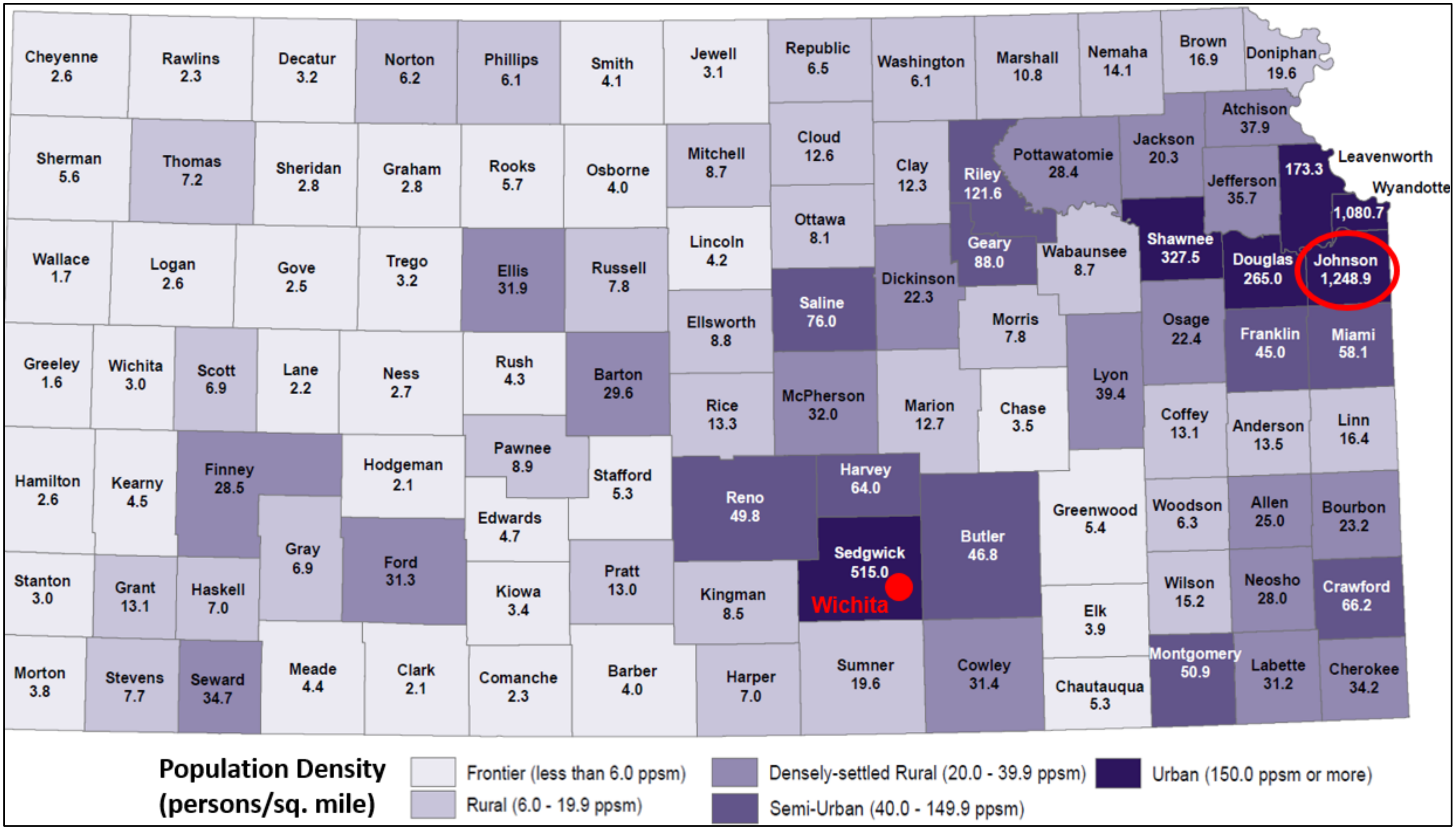


Figure 20. Population Density in Kansas by County (IPSR, 2017)

Table 2. The Population of the 35 Most Populous Counties in Kansas (U.S. Census Bureau, 2019a)

Rank	County Name	2017 Estimated Population
1	Johnson	578,797
2	Sedgwick	510,484
3	Shawnee	178,392
4	Wyandotte	163,227
5	Douglas	117,806
6	Leavenworth	79,359
7	Riley	75,696
8	Butler	66,260
9	Reno	63,360
10	Saline	55,334
11	Crawford	39,099
12	Finney	37,097
13	Geary	35,796
14	Cowley	35,732
15	Harvey	34,683
16	Ford	34,658
17	Montgomery	33,463
18	Lyon	33,302
19	Miami	32,976
20	Ellis	28,877
21	McPherson	28,792
22	Barton	27,067
23	Franklin	25,599
24	Sumner	23,336
25	Pottawatomie	23,188
26	Seward	22,948
27	Labette	20,553
28	Cherokee	20,501
29	Dickinson	19,162
30	Jefferson	18,856
31	Atchison	16,466
32	Neosho	16,209
33	Osage	15,894
34	Bourbon	14,757
35	Jackson	13,322

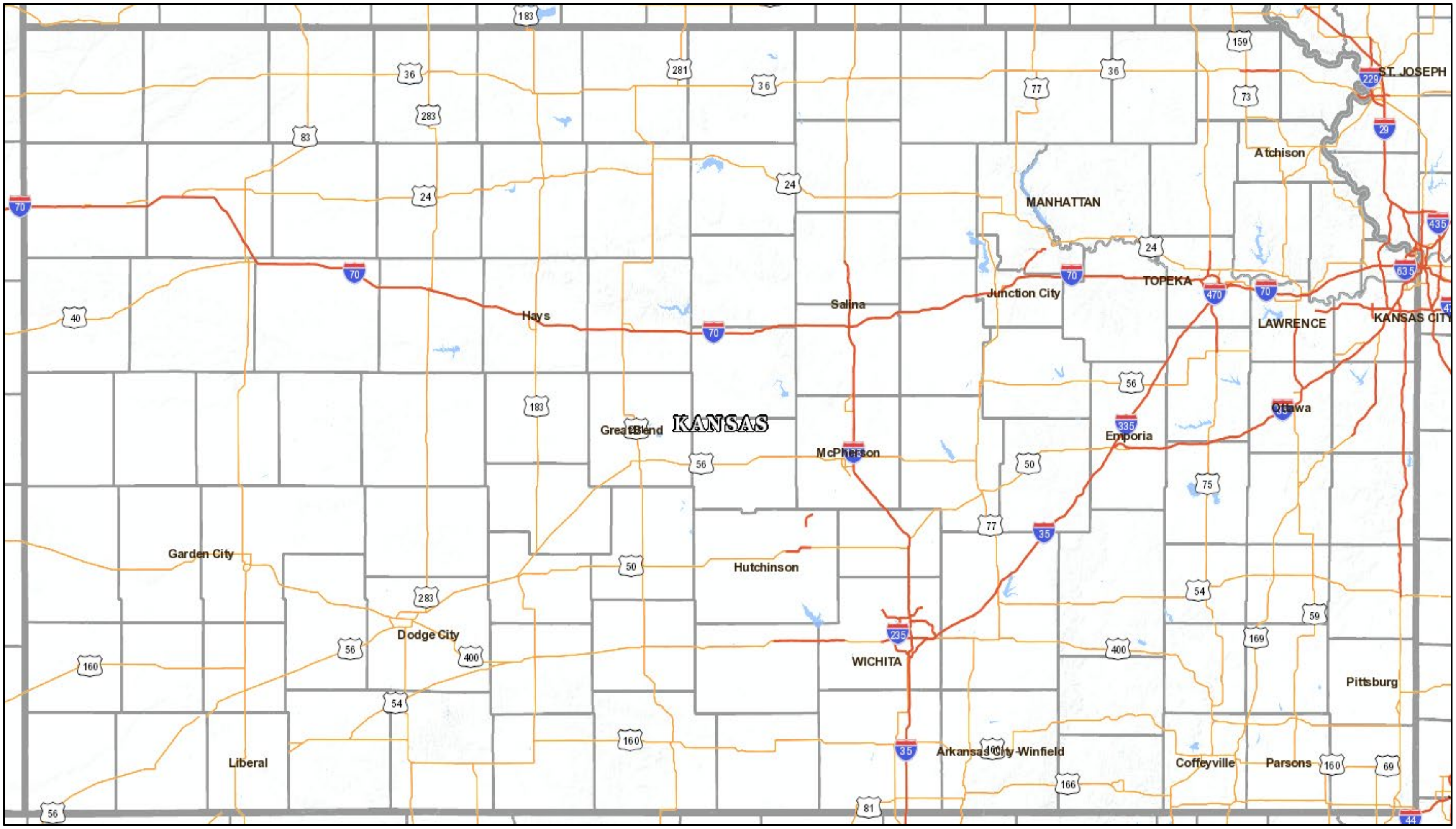


Figure 21. Major Interstate Highways in Kansas (U.S. Census Bureau, 2017)

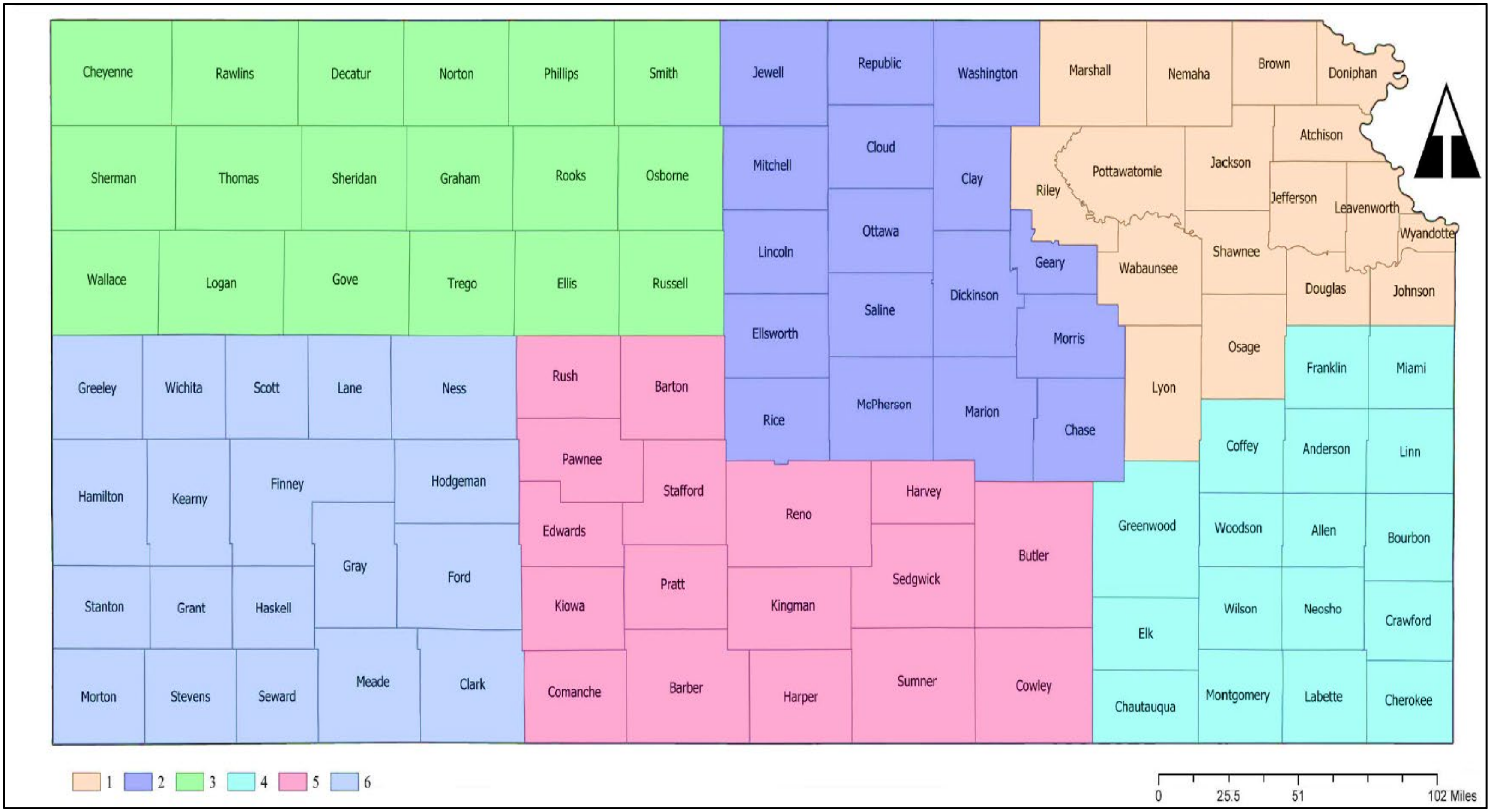


Figure 22. Kansas District Boundaries

- **CLASS E** mostly includes routes used for local services that only carry very short trips. Class E routes are frequently used daily to connect rural residents with higher speed routes. Examples for this class include K-76 and K-245, which carry about 800 vehicles daily on average.

Besides the State of Kansas, Wichita Unified School District (USD 259) also was selected as an example of a small study area to conduct some of the major functions of the descriptive and spatial analyses. USD 259 covers most of the city of Wichita, and it is one of the most populous districts in the state.

Teens aged 15-19 years old represent 6.9 percent of the Kansas population (U.S. Census Bureau, 2019a). Almost all of the teens in Kansas are distributed among 382 high schools across the 286 unified school districts shown in Figure 23, and 89 higher education centers shown in APPENDIX A (Table 29) (Kansas State Department of Education, 2019; IPEDS, 2019). Teen drivers in Kansas are an overrepresented group in motor vehicle crashes in comparison to other age groups; even though they represent approximately seven percent of the registered drivers in the state, they accounted for 19 percent of all traffic crashes (KDOT, 2015a).

KDOT has allocated a specific team (Teen Driver Emphasis Team) to develop and monitor a research-based action plan according to particular goals, as part of the SHSP, that could reduce the number and severity of crashes involving teen drivers (KDOT, 2015a).

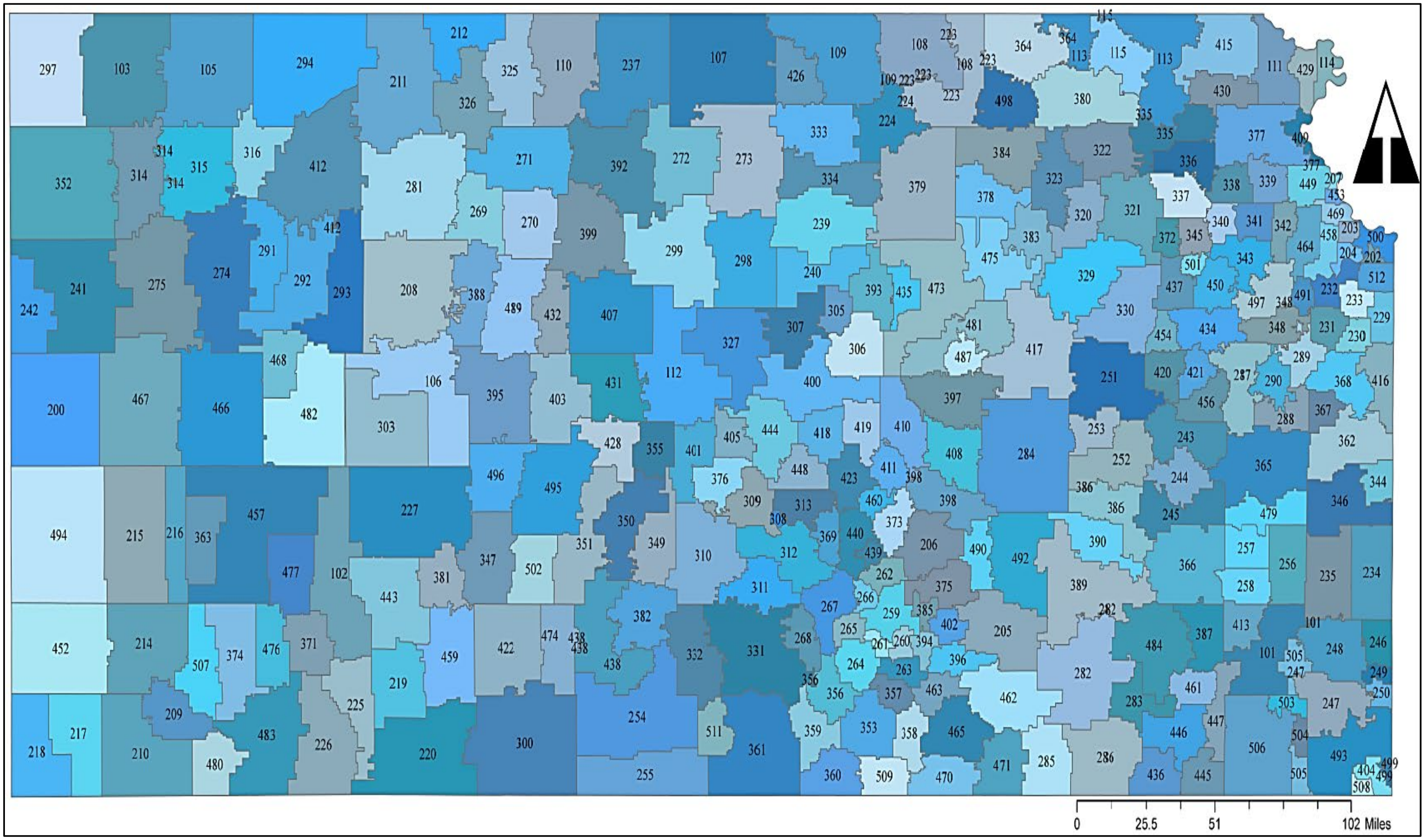


Figure 23. Kansas Unified School District Boundaries

This research is an attempt to identify potential factors that affect the number of crashes involving teen driver in order to conduct appropriate research-based actions to improve the traffic safety of the targeted drivers

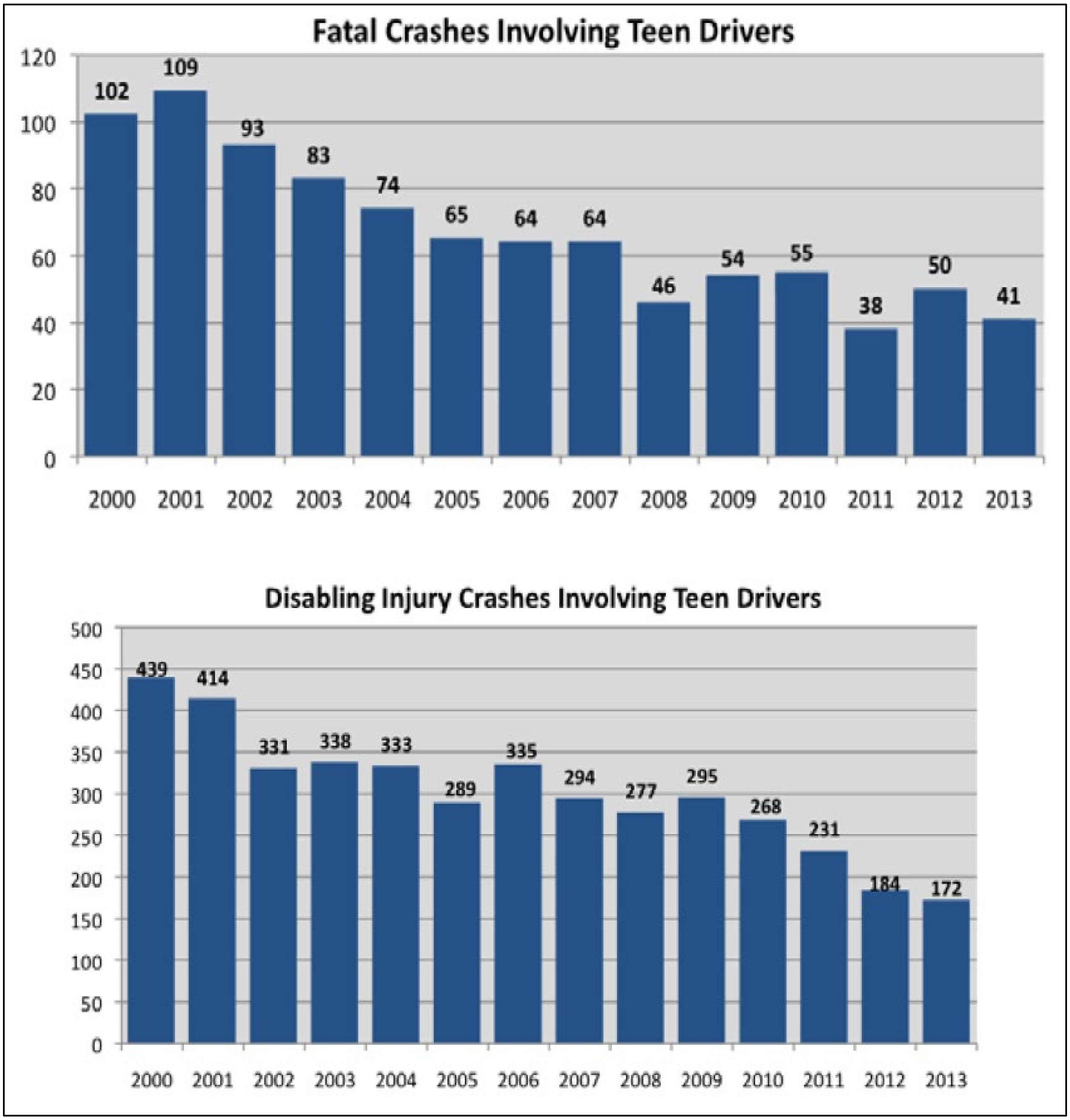


Figure 24. Fatal and Disabling Crashes Involving Teen Drivers (2000-2013) (KDOT, 2015a)

DATA COLLECTION AND PRE-PROCESSING

The required datasets for this study were broad and linked to different parties because each party had a piece related to analyzing traffic crashes involving teen-related crashes. Therefore, the major datasets used in this research were gathered from several resources such as the Fatality Analysis Reporting System (FARS) database, Kansas Department of Transportation database, U.S. Census Bureau Database, and Kansas Department of Education database.

FARS Database

FARS is a nationwide open-access database, which is maintained by the National Highway Traffic Safety Administration (NHTSA) under the U.S. Department of Transportation (USDOT). It includes yearly data of fatal injuries occurred in motor vehicle traffic crashes across the United State (NHTSA, 2019). The provided data include various parameters and details that were reported in the crash investigation processes and usually reported yearly from state DOTs.

The database contains information on fatal crashes from police reports, and it has been active since 1975. The required data are obtained online using the FARS Query System in six straightforward steps, as shown in the illustrative example in APPENDIX B. The data could be queried and downloaded in different formats for each year separately and the entire United States. Herein the .dbf files were utilized to download fatal crashes involving teen drivers in Kansas from 2010 to 2016. In this research, the obtained data from FARS were used in the fatal crash analysis.

KDOT Database

Part of the KDOT database includes all motor vehicle crashes on Kansas roadways, which were maintained by The Crash Data Unit stationed under the Bureau of Transportation Safety and

Technology at KDOT headquarters in Topeka. The Crash Data Unit is the primary repository in the state for records of all motor vehicle crashes that occurred from 2000 to the time of this research. This database is a limited-access resource for the public. The public has access to most records maintained by public entities and has an option to request more information such as maps, agency contracts, accident reports, and design plans from KDOT divisions and bureaus like the Bureau of Right of Way, Traffic Safety Division, and General Requests Division (KDOT, 2019). The data obtained from the KDOT database included all fatal and nonfatal traffic crashes with much more details and attributes than what FARS database provided.

U.S. Census Bureau Database

The U.S. Census Bureau is a principal agency of the US Federal Statistical System that produces and manage data related to the U.S. population and economy. They provide a free open-access database for public use. The main purpose of this database in this research was for collecting demographic characteristics of the state and its counties.

Kansas State Department of Education Database

The Kansas State Department of Education Database is a limited-access database run by the Kansas Department of Education. The database provides aggregate performance and demographic data on counties, districts, and schools in the State of Kansas. It also provides some GIS data on Unified School Districts (USD) boundaries, the number of schools, and the type of schools in Kansas. Hence, it became a part of the data resources that were used in this research.

Spatial Layers

To use the collected data from the sources mentioned above in spatial analysis, the data were converted from Excel format using coordinates (latitude and longitude) of each feature to a point shapefile in the ArcGIS software. For instance, the data obtained from FARS, which contain information on each fatal crash involving teen drivers, including coordinates of the crash location, were uploaded to the ArcGIS software using Display X, Y Data tool and converted into a point shapefile and represented geographically as a spatial layer on a map. This procedure was also applied to other datasets obtained from KDOT database, U.S. Census Bureau Database, and the Kansas Department of Education database to convert nonfatal crashes, demographic information of the state, and the location of educational centers to shapefiles, respectively.

In addition to collecting some existing GIS layers, for instance, state and county boundary shapefiles, these type of files are provided from various open resources and in different levels: global levels such as Esri, national levels such as USGS, state levels such as KDOT, and local levels such as counties and cities.

For the purpose of this research, a mini-spatial database was created, which contained different spatial layers related to teen drivers such as road networks, counties, high schools, and other educational institution zones, horizontal curves, etc., in Kansas. Additionally, this database included generating new layers and join the information (in csv. format) of teen driver-related crashes into existing layers, which were necessitated in the spatial analysis processes.

Data Reduction

The collected data were evaluated and arranged through the data reduction process to be prepared for the analysis process, but different challenges were confronted throughout this stage. For instance, as the injury scales that describe the injury severity of crashes from both of the resources were different, the severity was unified, based on available information in the databases. There were two major scales to rank the injury severity in traffic crashes, the Abbreviated Injury Scale (Hakkert and Braimaister) and the KABCO scale.

The AIS is an anatomical scoring system first introduced by the Association for the Advancement of Automotive Medicine in 1969, which codes individual injuries on a scale of 1 to 6, with one being minor and 6 represent fatal (Wong and Kunz, 2017). The KABCO scale was introduced by the National Safety Council in the late 1960s and codes injury severity on five levels, where K injuries are fatal, A injuries are incapacitating, B injuries are non-incapacitating, C injuries are possible injuries or complaint of pain, and the O injury severity is not injured (Compton, 2005). Herein, the KABCO scale was adapted in the analysis.

Another challenge during data reduction procedures was related to incomplete data. Incomplete data herein represent the data of crashes that have some blank fields. For instance, for the fatal crash analysis, there was a lack of coordinates for some of the crash cases that are recorded in the KDOT database. But there was no such problem in the FARS database because all recorded fatal crashes had coordinates. After evaluating the number of such cases, it was found that those crashes were distributed randomly across the state and did not have any patterns. Therefore, these uncompleted crash data were used in the descriptive analysis phase, but they were removed in the spatial analysis phase.

Table 3 shows, by category, the number of crash cases removed from the total number of crashes involving teen drivers for each specific analysis process and the reason of the removal, which generally displays missing information of the omitted cases. Therefore, the number of features reported in the database might not match the number used in the analysis. The categories are organized in the order of discussion in this research.

Table 3. Number of Crash Cases Reduced from Analysis

Category	Total No.	Removed Cases	Used Cases	Reason for Removal
Gender	82,564	69	82,495	Reported as Unknown
Seat Belt Usage	82,564	586	81,978	Left Blank
Injury Severity	82,564	980	81,584	Reported as Unknown
Wearing Seatbelt	82,564	586	81,978	Left Blank
Safety Equipment Use	82,564	3,292	79,272	Reported as Unknown/ Misreported[1]
Vehicle Body Type	82,564	35	82,529	Reported as Unknown
Vehicle Year	82,564	360	82,204	Left Blank/ Misreported[2]
Crash Location	76,191	19	76,172	Reported as Unknown
Non-State Road Function Class	76,191	19,459	56,732	Left Blank
Weather Conditions	76,191	191	76,000	Reported as Unknown
Light Conditions	76,191	144	76,047	Reported as Unknown
Date	76,191	133	76,058	Left Blank
Months	76,191	134	76,057	Left Blank
First Harmful Event	76,191	74	76,117	Reported as Unknown
Collision with Other Vehicles	52,204	103	52,101	Reported as Unknown
Coordinates (All)[3]	76,191	3,535	72,656	Left Blank
Coordinates (Non-fatal)[4]	72,370	286	72,565	Fatal Crashes

[1] In some cases, the safety equipment used were reported as child seats. Therefore, they were removed from the dataset.
[2] Some vehicles' model year were listed as 20133, 2099, 2100. These vehicles were removed from the dataset.
[3] This includes data of all types of crashes (fatal, injury, and PDO) involving teen drivers, downloaded from the KDOT crash database.
[4] After excluding fatal crashes from all crashes, the number of non-fatal crashes was 72,370. For spatial analysis of fatal crashes, the FARS database was used.

CHAPTER V. RESEARCH ANALYSIS

DESCRIPTIVE ANALYSIS

The teenage group (15- to 19-years old) represents seven percent of the Kansas population on average, as shown in the age distributions for the years between 2010 and 2016, as shown in APPENDIX B, Table 30. Of the average percentage of teenagers, 51 to 52 percent were males aged 16-19 and 48 to 49 percent were females (see Table 4). The number of teenagers aged 19 is higher than the others.

Table 4. Average Population and Percentage of Teenagers

Age\Gender	Male	Male (%)	Female	Female (%)	Both genders
16	20,377	0.51	19,250	0.49	39,627
17	20,406	0.51	19,235	0.49	39,641
18	20,423	0.52	19,134	0.48	39,557
19	22,291	0.52	20,309	0.48	42,599

The number of licensed teen drivers in Kansas, including restricted drivers and graduated driver licenses is shown in Table 5. The table identifies the age and gender distribution of teen drivers for 2010 to 2016 and shows that the number of drivers aged 16-19 increased with age linearly for both genders aged 16-19 — that is, the number of licensed teen drivers aged 19 is higher than other ages.

Table 5. Licensed Teen Drivers In Kansas, By Age and Gender

Year	Gender	<16	16	17	18	19
2010	Male	17,520	13,212	15,561	17,177	17,682
	Female	16,371	12,601	14,860	16,496	17,283
	Total	33,891	25,813	30,421	33,673	34,965
2011	Male	13,730	12,701	15,069	16,728	17,653
	Female	12,926	12,280	14,442	15,831	17,051
	Total	26,656	24,981	29,511	32,559	34,704
2012	Male	13,680	12,116	14,655	16,450	17,318
	Female	12,945	11,524	14,207	15,726	16,396
	Total	26,625	23,640	28,862	32,176	33,714
2013	Male	14,269	12,133	14,390	16,272	17,341
	Female	13,545	11,488	13,800	15,748	16,517
	Total	27,814	23,621	28,190	32,020	33,858
2014	Male	14,257	12,582	14,301	16,046	17,127
	Female	13,417	11,971	13,635	15,342	16,464
	Total	27,674	24,553	27,936	31,388	33,591
2015	Male	14,321	12,642	14,730	16,049	16,909
	Female	13,793	11,712	14,118	15,235	16,205
	Total	28,114	24,354	28,848	31,284	33,114
2016	Male	13,835	12,446	14,699	16,411	16,820
	Female	13,518	11,865	13,696	15,553	15,907
	Total	27,353	24,311	28,395	31,964	32,727
Average	Male	14,516	12,547	14,772	16,448	17,264
	Female	13,788	11,920	14,108	15,704	16,546
	Total	28,304	24,468	28,880	32,152	33,810

The average Kansas resident in the teenage group in both genders who held a driver license is shown in Table 6 and indicates that even though the number of male residents and licensed male teens were higher than female residents and licensed female teens (see Table 4 and Table 5), the licensure rate of females was greater than males aged 16-19. This comparison is visualized in Figure 25.

Table 6. Average Teen Driver Licenses, Kansas Residents, and Licensure Rate by Age and Gender

Age		<16	16	17	18	19
Male	Residents (thousands)	329.44	20.38	20.41	20.42	22.29
	Licenses (thousands)	14.52	12.55	14.77	16.45	17.26
	Licenses per 100 residents	4.41	61.58	72.39	80.53	77.45
Female	Residents (thousands)	314.23	19.25	19.24	19.13	20.31
	Licenses (thousands)	13.79	11.92	14.11	15.70	16.55
	Licenses per 100 residents	4.39	61.92	73.35	82.08	81.47
Both genders	Residents (thousands)	643.67	39.63	39.64	39.56	42.60
	Licenses (thousands)	28.30	24.47	28.88	32.15	33.81
	Licenses per 100 residents	4.40	61.74	72.86	81.28	79.37

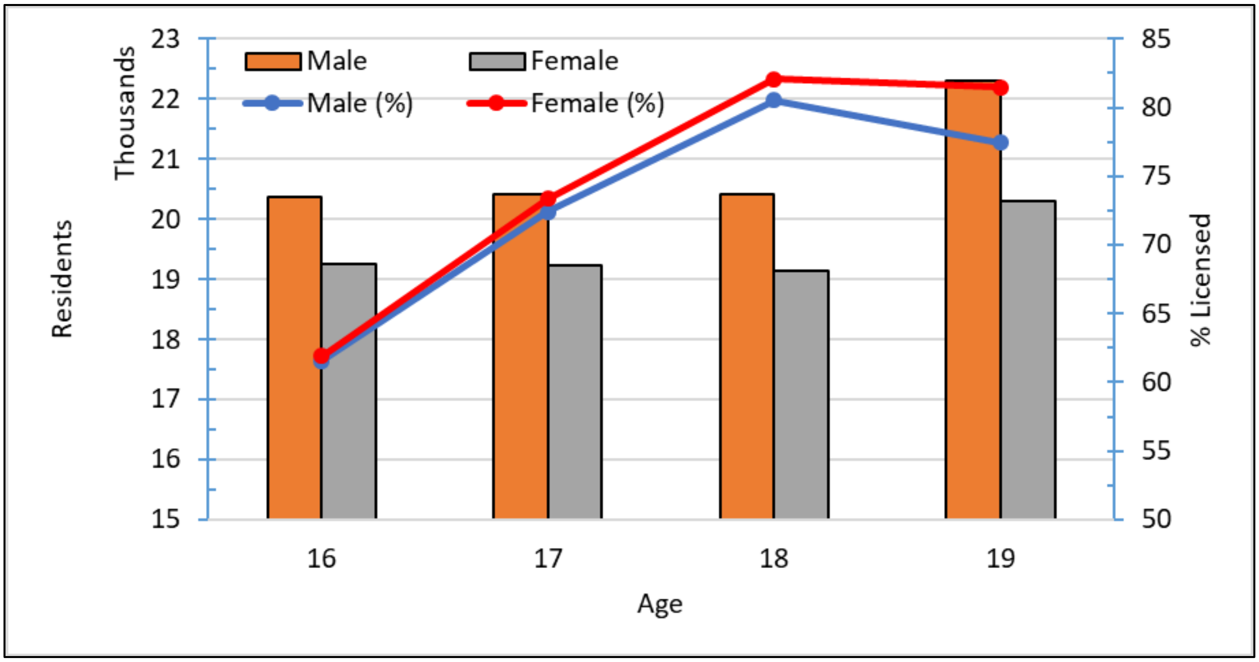


Figure 25. Population and Licensed Teenagers by Age and Gender

In this research, the traffic crash dataset involving crashes by teen drivers were used to categorize and analyze different characteristics associated with crashes and correlation between characteristics. These characteristics are listed under three related factors: crash, driver and vehicle.

Crash Characteristics

This section comprises characters of traffic crashes involving teen drivers that occurred between 2010 and 2016. These characters include crash conditions, type of crashes, and crash location. The crash data were downloaded from the KDOT database after applying a filter. This filter was carried out on three levels — driver indicator, teen driver indicator, and driver age between 15 and 19.

To ensure the data were not duplicated (same crash, driver, or vehicle counted twice or more), the information on crash ID, time and date of crashes, drivers' full names and ages involved in crashes, and vehicles involved in crashes were downloaded. When duplicated crashes were found, the drivers' names and vehicle types were compared, and then duplicated drivers or crashes were deleted. The total number of crashes for each year, including all drivers and teen drivers, were calculated. The results showed that the total crashes occurred in Kansas in the six years was 422,238 crashes, and the number of crashes involving teen drivers was 76,191 crashes, as shown in Table 7.

Table 7. Crashes Involving Teen Drivers by Year

Year	Total Crashes	Crashes Involving Teen Drivers	Crashes Involving Teen Drivers (%)
2010	60,562	11,765	19.43
2011	60,270	10,978	18.21
2012	58,373	10,587	18.14
2013	59,265	10,472	17.67
2014	59,938	10,508	17.53
2015	61,440	10,709	17.43
2016	62,390	11,172	17.91
Total	422,238	76,191	18.04

Of the 76,191 crashes, teen drivers involved in 300 fatal crashes. In 2011, the lowest number of fatal crashes involving teen drivers was reported while the lowest number of crashes was in 2013. However, the greatest number of all type of severity crashes was in 2010, as shown in Table 8. The annual average number of crashes involving teen drivers for the study period was 10,884 crash, and the annual average number of fatal crashes was 43 crashes. The results imply that the number of crashes involving teen drivers is in a considerable fluctuation pattern instead of the targeted linear decline pattern.

Table 8. The Severity of Crashes Involving Teen Drivers.

Crash Severity	2010	2011	2012	2013	2014	2015	2016	Total
Fatal	54	37	49	39	41	38	42	300
Injury	2,975	2,802	2,764	2,580	2,573	2,601	2,632	18,927
PDO	8,736	8,139	7,774	7,853	7,894	8,070	8,498	56,964
Total	11,765	10,978	10,587	10,472	10,508	10,709	11,172	76,191

Crashes in Counties and KDOT Districts

In the 105 counties of Kansas, Johnson County, which is the most populous county in the state, had the highest number of crashes involving teen drivers. However, the greatest number of fatal and injury crashes were in Sedgwick County, the second most populous county in the state, as shown in Figure 26. The distribution of the severity of crashes, especially fatal crashes, were varied among counties. For instance, although Douglas County is the fifth most populous county, it had the lowest number of fatal crashes when compared to the 30 counties that had a high number of crashes. Conversely, Leavenworth County had the largest number of fatal crashes when compared to the 30 counties that had a high number of crashes.

Also, both Johnson County and Sedgwick County had the highest numbers of crashes involving teen drivers in 2016, which were 2,341 and 2,379 crashes, respectively. Similarly, when that data were analyzed based on the KDOT districts, District One had the greatest number of crashes and District Five in the secondmost. Expectedly, District Three had the lowest number of crashes.

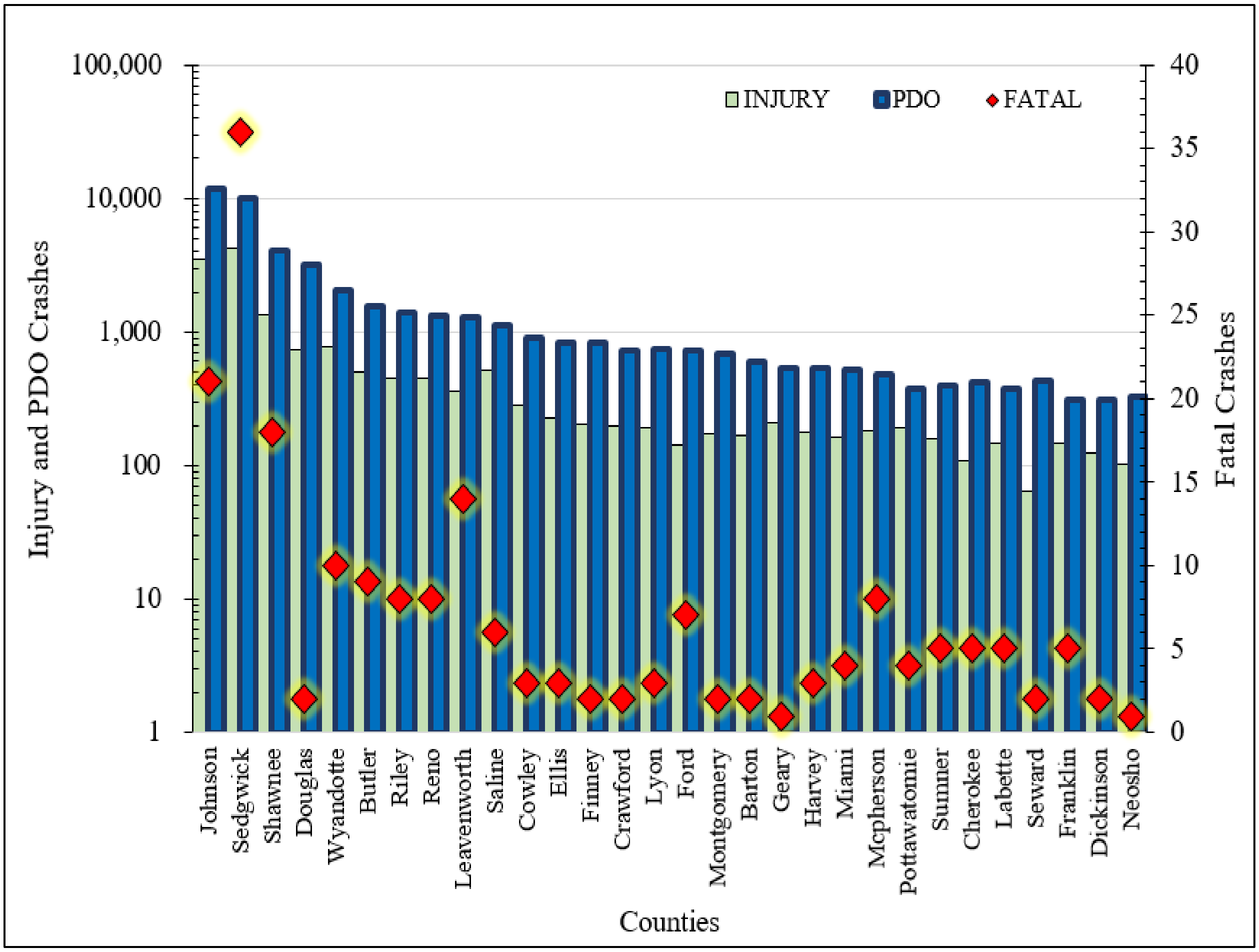


Figure 26. The 30 Counties That Had the Largest Number of Crashes Involving Teen Drivers.

Crashes in Intersections and on Horizontal Curves

The results showed that 40.74 percent of crashes involving teen drivers were intersection-related crashes and only 7.84 percent of the crashes were on horizontal curves (see Table 9 and Table 10). Of the 300 fatal crashes involving teen drivers, Teen drivers were involved in 25.33 percent of intersection-related fatal crashes and 15.33 percent of fatal crashes that occurred on horizontal curves. This indicates that intersections represent considerably more critical locations for teen drivers than horizontal curves.

Table 9. Crash Severity of Teen Drivers in Intersections and on Horizontal Curves

Crash Location	Intersection-Related	Non-Intersection Related	On Horizontal Curves	Off Horizontal Curves
Fatal	76	224	46	254
Injury	8,222	10,705	1,632	17,295
PDO	22,744	34,220	4,292	52,672
Total	31,042	45,149	5,970	70,221

Table 10. Percentage of Crash in Intersections and on Horizontal Curves by Severity

Crash Location	Intersection-Related	Non-Intersection Related	On Horizontal Curves	Off Horizontal Curves
Fatal	25.33	74.67	15.33	84.67
Injury	43.44	56.56	8.62	91.38
PDO	39.93	60.07	7.53	92.47
Total	40.74	59.26	7.84	92.16

Crashes on Local Roads and State Highways

The analysis was conducted on crashes involving teen drivers that occurred on local roads and state highways. The local roadways include roads under the local's jurisdiction while the state highways are those under the state's jurisdiction and mostly located outside of city boundaries. State route names contain an "I," "US," or "K" prefix in Kansas (KDOT, 2008). Even though the local roads in the state carry only 45 percent of the traffic, the results showed that they had 74.46 percent of all crash severity types that involved teen drivers when compared with the state highway system, as shown in Table 11.

Table 11. The Severity of Crashes Involving Teen Drivers on Local Roadways

Roadway Type	Local Roadway	State Highway System
Fatal	171	129
Injury	14,030	4,897
PDO	42,531	14,433
Total	56,732	19,459

Table 12. Percentage of Crashes on Local Roadways by Severity

Roadway Type	Local Roadway	State Highway System
Fatal	57.00	43.00
Injury	74.13	25.87
PDO	74.66	25.34
Total	74.46	25.54

With reference to functional classification, the greatest portion of the crashes involving teen drivers was on local roads (34.47 Percent) and arterials, as principal arterials and minor arterials, were in the second level (see APPENDIX B, Table 37). Although the speed limit on local roads is low in comparison to other classes, 46.78 percent of fatal crashes that occurred on the local roadway network were in the local class.

Crashes and Weather Conditions

The number of crashes involving teen drivers in relation to weather conditions at the crash time is shown in Figure 27. The results indicate that the majority of crashes involving teen drivers occurred in clear weather conditions. The lowest number of fatal crashes involving teen drivers was on snowy days; however, the injury and PDO crashes on those days were greater than the number of crashes occurred on foggy and windy days.

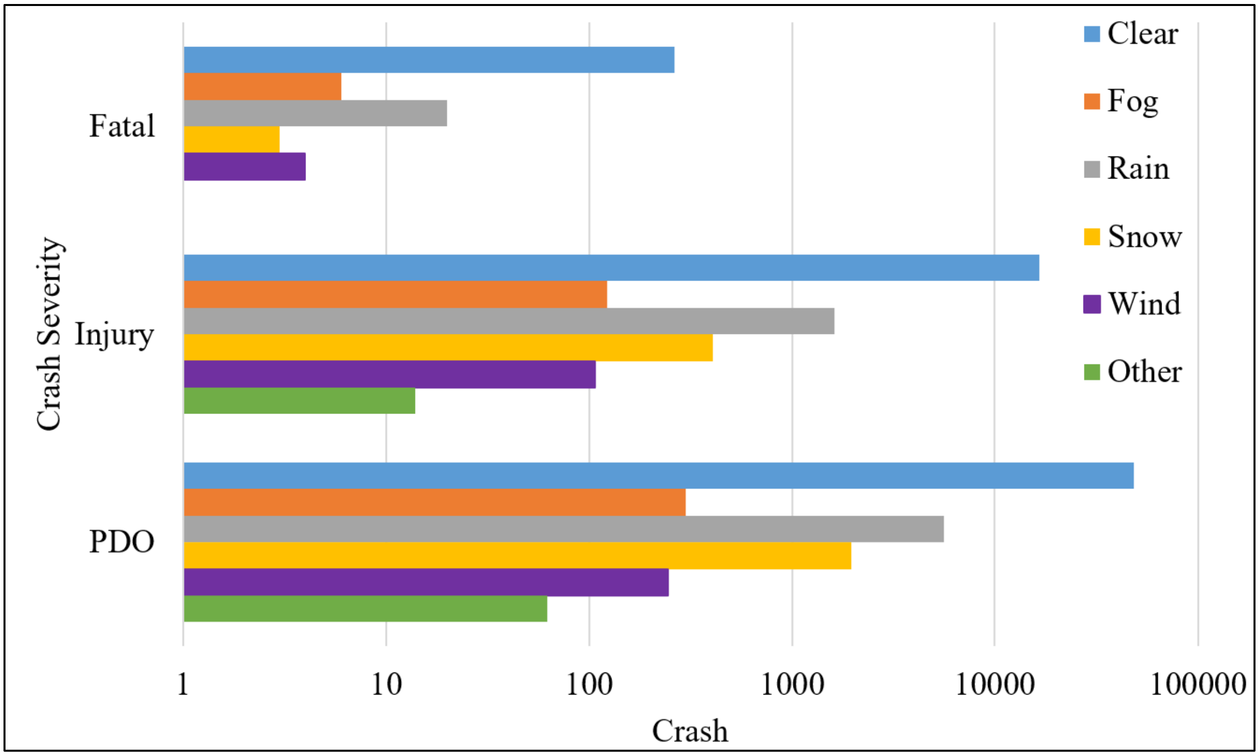


Figure 27. The Severity of Teen-related Crashes in Different Weather Conditions

Crashes and Light Conditions

The results are shown in Figure 28 representing the number of crashes in different light conditions. More than 26 percent of all crashes involving teen drivers were in dark conditions while 40.5 percent of the fatal crashes were in dark conditions. Of the fatal crashes that occurred in dark conditions, 73.55 percent of them were in the absence of street lights. At first glance, all crash severity types were higher during daylight, but this statement is inaccurate for the fatal crashes. The average duration of each dawn and dusk is about 30 minutes while the duration of daylight and darkness is about 11.5 hours (Time and Date, 2019).

By normalizing[1] the number of crashes involving teen drivers that occurred during dawn and dusk into pre-hour rates, the results showed that the rate of fatal crashes at those periods jointly was 22.71 percent greater than during daylight and 41.55 percent greater than during dark conditions. Comparing crash rates in the dawn and dusk periods to other periods (day and night) show that all types of crash rates were higher during dawn and dusk by more than 13 percent and more than 32 percent of fatal crashes. The results of this analysis are shown in APPENDIX B, Table 39.

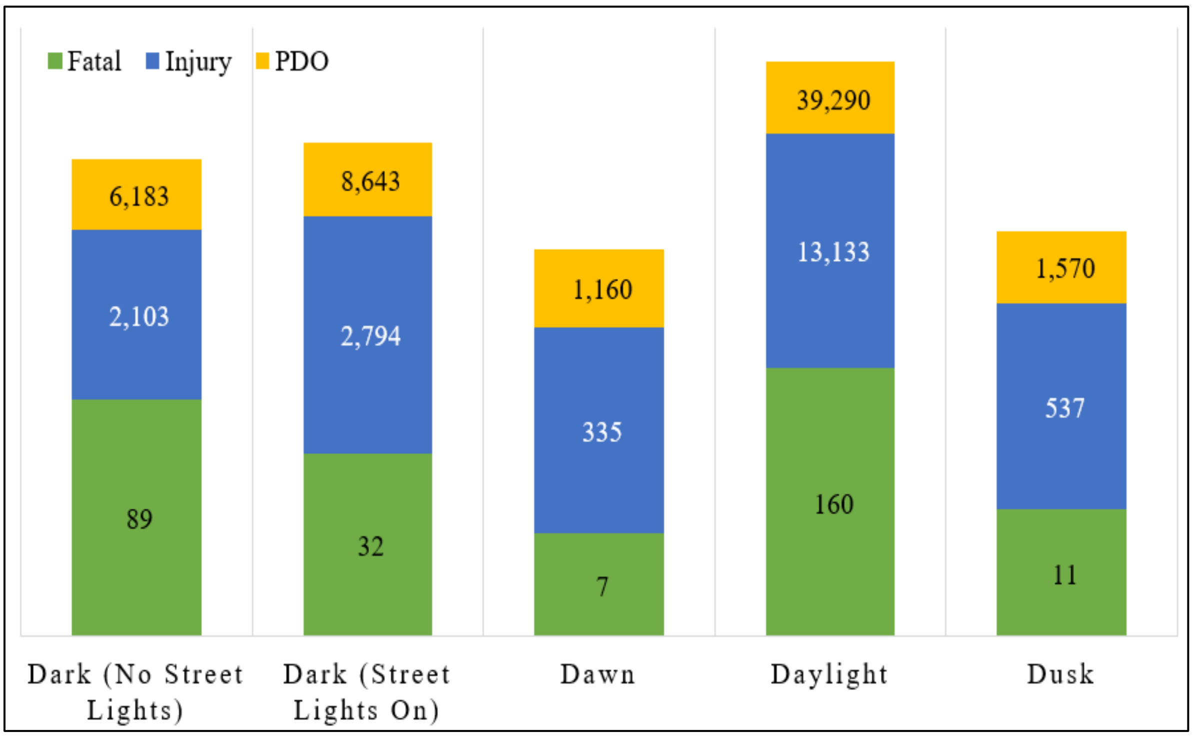


Figure 28. The Severity of Crashes by Light Conditions

[1] A new measure was, N, that effectively normalizes for the prevalence of the light conditions. This is defined for dark condition as: $N_K = K/11.5$ where K is number of crashes during dark condition and 11.5 represent number of hours of dark condition. Similarly for Daylight condition, $N_D = K/11.5$ where D is number of crashes during daylight condition and 11.5 represent number of hours of daylight condition. Since duration of dawn and dusk is one hour, normalizing was not needed.

Crashes by Time, Days, and Months

Breaking down the crashes by each hour of the day shows the time of day that experiences the highest number of crashes. The results displayed in Figure 29 show that the distribution of the crashes involving teen drivers reflects the group's daily activity pattern. The time in the figure begins at midnight, which represents the time from 12:00 a.m. to 12:59 a.m. The morning peak hour was at 7:00 a.m. (7:00 a.m.-7:59 a.m.) when most of the schools and colleges start. Most of the crashes occurred at 8:00 a.m. was on Wednesdays, when schools start late (typically 8:50 a.m.). The afternoon peak hour was at 3:00 p.m. (3:00 p.m.-3:59 p.m.), which constitutes the highest part of the day when high schools usually are dismissed in Kansas. Other spikes were at noon, which is the lunchtime and at 5:00 p.m. during the evening traffic peak hour.

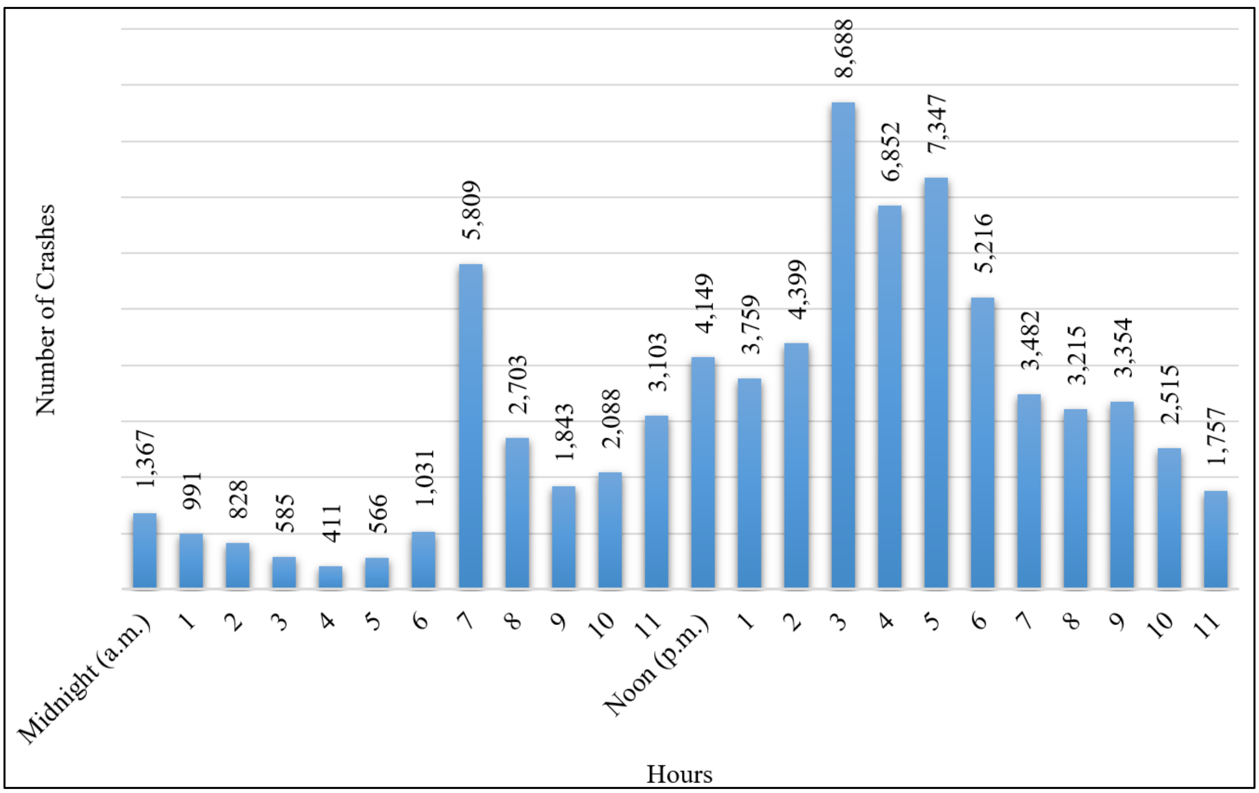


Figure 29. The Pattern of Crashes Involving Teen Drivers by Hours

The pie chart shown in Figure 30 demonstrates the proportion of crashes involving teen drivers throughout the days of the week. The weekends in general, and Sundays especially, experienced the least number of the crashes. However, the weekends and Sundays specifically experienced the greatest number of fatal crashes (see APPENDIX B, Table 40). Most crashes happened on weekdays and Fridays received the highest share. The results indicated that most of the crashes happened on Fridays in the afternoon hours. Moreover, a large number of crashes were recorded in the early hours of the day (12:00 a.m. - 5:00 a.m.) on weekends. The details of these results are presented in APPENDIX B, Figure 59.

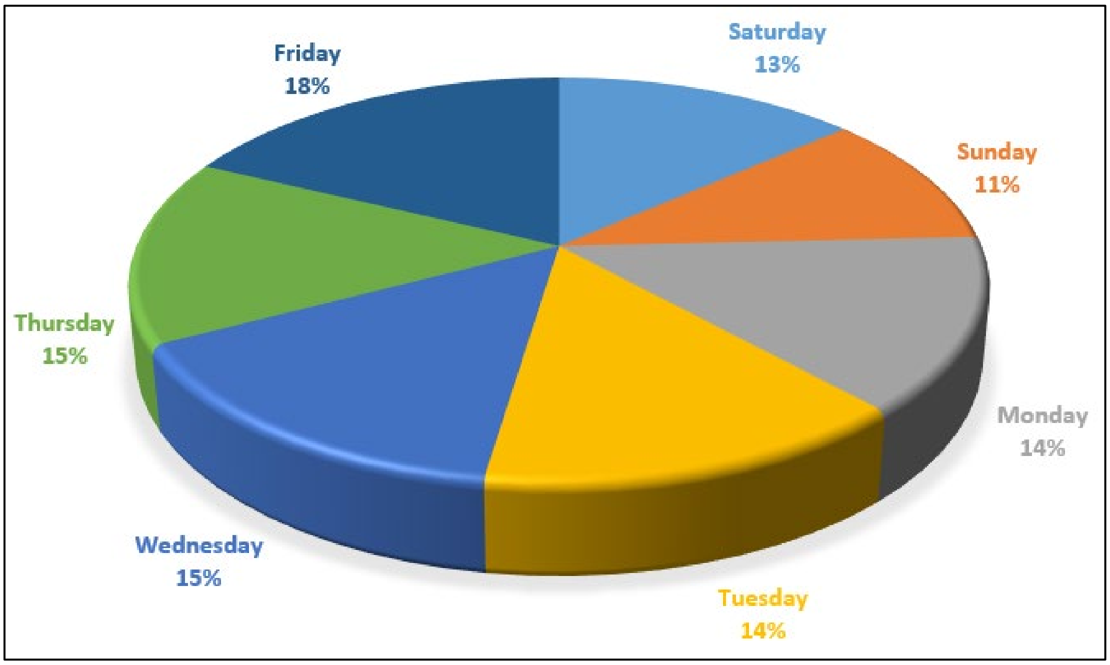


Figure 30. Crashes Involving Teen Drivers by the Days of the Week

As regards the breakdown of overall crashes involving teen drivers based on months, October experienced the greatest number of crashes of overall crash severity and each type of crash severity. Conversely, March experienced the least number of crashes involving teen drivers, as shown in Figure 31.

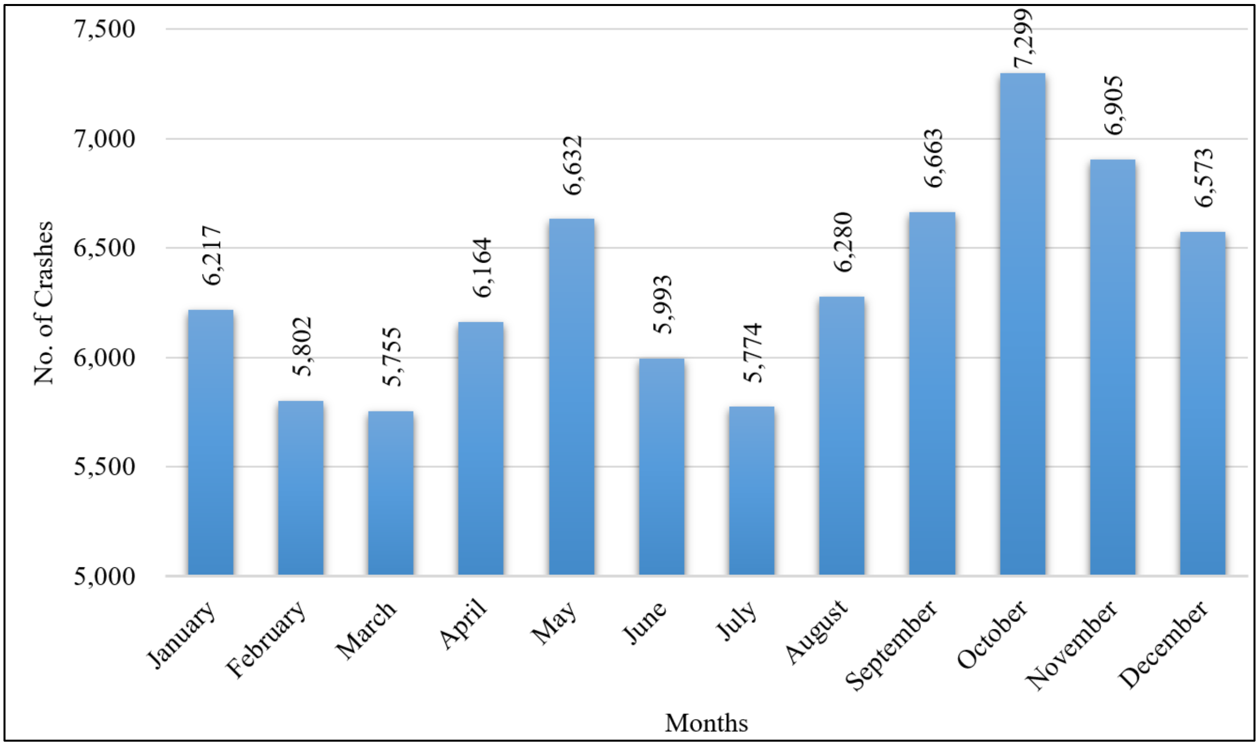


Figure 31. Number of Crashes by Month

Crash Types

When analyses were conducted based on classifying crashes between multi-vehicle crashes, as also known as Collision with Other Vehicle (CWOV), and single-vehicle crashes, the results showed that only 31.4 percent of overall crashes were single-vehicle crashes, but constituted 52.5 percent of the fatal crashes. By breaking down the single-vehicle crashes, crashes with fixed objects[1] constituted 20.9 percent of overall crashes and 31.8 percent of fatal crashes, as shown in Table 13. This means the single-vehicle crashes were more severe than CWOV for teen drivers.

The 2,648 overturned crashes shown in Table 13 represents only 3.5 percent of the whole number of crashes, but 13.4 percent of fatal crashes. In term of the location of these overturned

[1] Collision with fixed objects included any fixed objects, collision with parked vehicles, or dead bodies (animal or human).

crashes, about 86.6 percent of them happened on rural roads, but this percentage was much higher for fatal crashes. For instance, the 40 fatal crashes occurred in the overturned crashes (see Table 13), 39 of them (97.5 percent) were on rural roads.

Table 13. Crashes Involving Teen Drivers by Crash Type

Crash Type	Fatal		Injury		PDO		Total	
	No.	(%)	No.	(%)	No.	(%)	No.	(%)
Multi-vehicle	142	47.49	12,571	66.48	39,491	69.39	52,204	68.58
Single-vehicle	157	52.51	6,338	33.52	17,418	30.61	23,913	31.42
Fixed Object	95	31.77	4,003	21.17	11,778	20.70	15,876	20.86
Animal	1	0.33	175	0.93	3,716	6.53	3,892	5.11
Overturned	40	13.38	1,469	7.77	1,139	2.00	2,648	3.48
Other[1]	7	2.34	337	1.78	759	1.33	1,103	1.45
Pedestrian	10	3.34	222	1.17	8	0.01	240	0.32
Pedalcycle	3	1.00	130	0.69	10	0.02	143	0.19
Train	1	0.33	2	0.01	8	0.01	11	0.01

When it comes to CWOV, most of those crashes were rear end (by 44.2 percent) and angle (by 40.3 percent) crashes, as shown in Figure 32. However, the majority of fatal crashes in CWOVs were angle types (by 61.3 percent) and head on types (by 23.9 percent). The rear end crashes represented only 7.0 percent of the fatal crashes. This means, in contrast to the angle and head-on crashes, the number of rear-end crashes was highest, but their severity was lower. The results of the analysis are illustrated in APPENDIX B, Table 42.

[1] Other types include non-collision cases such as driving into water, striking holes or bumps, Jackknife, etc.

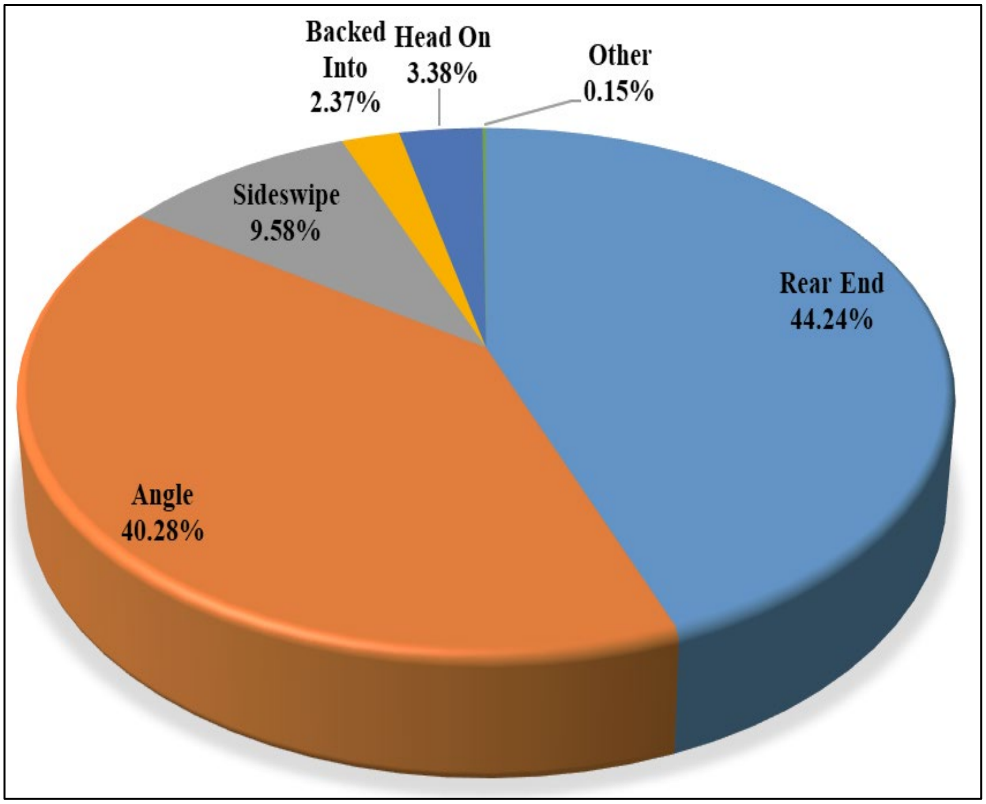


Figure 32. The proportion of Multi-vehicle Crashes Involving Teen Drivers

Driver Characteristics

This section comprises traits of teen drivers involved in traffic crashes including gender, age, and risky behavior such as DUI, seat belt usage, and distractions. Of crashes involving teen drivers, 7.7 percent (5,876 crashes) involved more than one teen driver. Therefore, there were 82,564 teen drivers involved in 76,191 crashes—this is, the majority of these crashes involved one teen driver and only 14.8 percent of teen drivers (12,249 teen drivers) were involved in crashes involving more than one teen driver. The results of these analyses are shown in Table 14.

Table 14. Number of Teen Drivers Involved in Crashes by Year

	Teen Drivers Involved in Crashes				Crashes Involving Teen Drivers			
Year /	1	2	>2	Total	1	2	>2	Total
2010	10,720	1,960	217	12,897	10,720	980	65	11,765
2011	10,169	1,506	183	11,858	10,169	753	56	10,978
2012	9,809	1,438	188	11,435	9,809	719	59	10,587
2013	9,674	1,496	163	11,333	9,674	748	50	10,472
2014	9,719	1,480	159	11,358	9,719	740	49	10,508
2015	9,935	1,440	170	11,545	9,935	720	54	10,709
2016	10,289	1,638	211	12,138	10,289	819	64	11,172
Total	70,315	10,958	1,291	82,564	70,315	5,479	397	76,191

The greater number of crashes involving teen drivers occurred in 2010 and 2016, which were 12,633 and 12,038 crashes, respectively. On the contrary, 2013 had the lowest number of teen drivers involved in crashes and the lowest number of teen driver fatalities, as shown in Table 15. The results showed different patterns in the number of crashes in term of severity. In looking only at the difference from 2010 to 2016, the overall number of crashes declined, but the amount was not the same for each severity type, as listed below:

- The fatal crash type dropped by 18.5 percent;
- The incapacitating crash type dropped 46.4 percent, which was the highest downward trend when compared to other severity types;
- The injury, not incapacitating type dropped by 16.8 percent;
- The possible injury type declined by 10.3 percent; and
- Lastly, for PDO crashes, there was not a lot of changes. It had the lowest downward trend (2.8 percent) in comparison to other severity types.

Table 15. Severity of Teen Driver Involved in Crashes by Year

Injury Severity	2010	2011	2012	2013	2014	2015	2016	Total
Fatal Injury (K)	27	20	21	16	23	21	22	150
Incapacitating (A)	151	119	104	98	82	83	81	718
Injury, Not Incapacitating (B)	752	720	712	649	635	679	626	4,773
Possible Injury (C)	909	804	841	790	787	793	815	5,739
Not Injured (O)	10,794	9,975	9,634	9,680	9,752	9,875	10,494	70,204
Total	12,633	11,638	11,312	11,233	11,279	11,451	12,038	81,584

The number and percentage of teen drivers involved in crashes in term of gender are shown in Table 16. Even though the numbers were fluctuating throughout the studied period, male teen drivers represent a greater portion of teen drivers involved in crashes, which constitute, on average, almost 53 percent of the total number of crashes involving teen drivers. However, this percentage changed vividly when the gender of teen drivers compared with the term of injury severity. Male teen drivers comprised more than 67 percent of fatalities and about 56 percent of disabilities, as shown in APPENDIX B, Table 41.

Table 16. Gender of Teed Drivers Involved in Crashes

Year	Female	Female (%)	Male	Male (%)	Total
2010	6,127	47.59	6,748	52.41	12,875
2011	5,496	46.40	6,349	53.60	11,845
2012	5,355	46.85	6,074	53.15	11,429
2013	5,323	47.00	6,003	53.00	11,326
2014	5,404	47.61	5,947	52.39	11,351
2015	5,424	47.03	6,110	52.97	11,534
2016	5,674	46.76	6,461	53.24	12,135
Sum/Avg.	38,803	47.04	43,692	52.96	82,495

In respect of the age of teen drivers involved in crashes, the results shown in Figure 33 indicate that the average number of teen drivers involving in crashes increased linearly with ages between 15-18 and dropped by 0.70 percent for drivers aged 19 (see APPENDIX B, Table 31 for more details). This means that teen drivers aged 18 were more likely to be involved in crashes than other teen drivers.

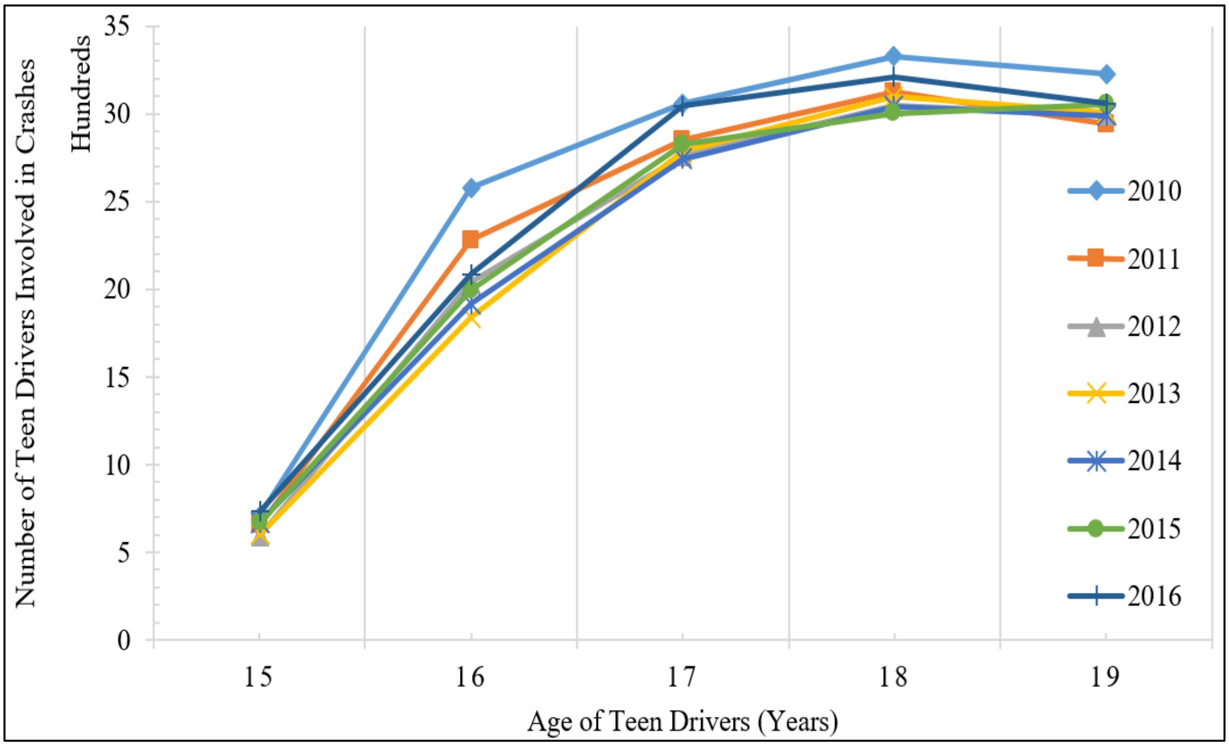


Figure 33. Age of Teen Drivers Involved in Crashes

The data showed that using seatbelts as a lapbelt, shoulder, or shoulder-lap and/or other safety equipment such as airbags, motorcycle helmets and/or eye protections had positive impacts on reducing injury severity of teen drivers involved in crashes (see Table 17). And, conversely, nonuse of none of this equipment dramatically increased injury severity of teen drivers involved in crashes. Most of teen driver fatalities (57.1 percent) occurred when teen drivers were unbelted.

Table 17. Safety Equipment and Seatbelts Impacts on Teen Drivers' Injury Severity

Injury Severity[1]	Safety Equipment				Seatbelt			
	Used	Used (%)	None Used	None Used (%)	Used	Used (%)	Not Used	Not Used (%)
K	86	60.56%	56	39.44%	60	42.86%	80	57.14%
A	492	74.32%	170	25.68%	402	61.28%	254	38.72%
B	3,965	86.76%	605	13.24%	3,722	81.48%	846	18.52%
C	5,181	93.96%	333	6.04%	5,060	89.75%	578	10.25%
O	66,949	98.73%	863	1.27%	66,727	95.31%	3,284	4.69%
Total	76,673		2,027		75,971		5,042	

The number of teen drivers involved in crashes under the influence of alcohol and/or drugs and the percentage of teen driver fatalities in total crashes involving teen drivers is shown in Figure 34. More than 27 percent, on average, of teen driver fatalities, were alcohol and/or drug-related. Further details are at APPENDIX B, Table 32 and Table 33.

Of 76,191 crashes involving teen drivers, drivers were distracted in 31.7 percent of them. The results shown in Figure 35 indicate that teen drivers were distracted in more than 29 percent of the crashes in 2010 and this proportion increased to about 34 percent in 2016. The distractions could be any activity that diverts drivers' attention (visually, manually, and/or cognitively) from safe driving. These activities including but not limited to using a cell phone, eating and drinking, talking to the occupants, and/or using other devices such as a radio or navigation system.

[1] Based on the KABCO classification: Fatal Injury (K), Incapacitating (A), Injury, Not Incapacitating (B), Possible Injury (C), and Not Injured (O)

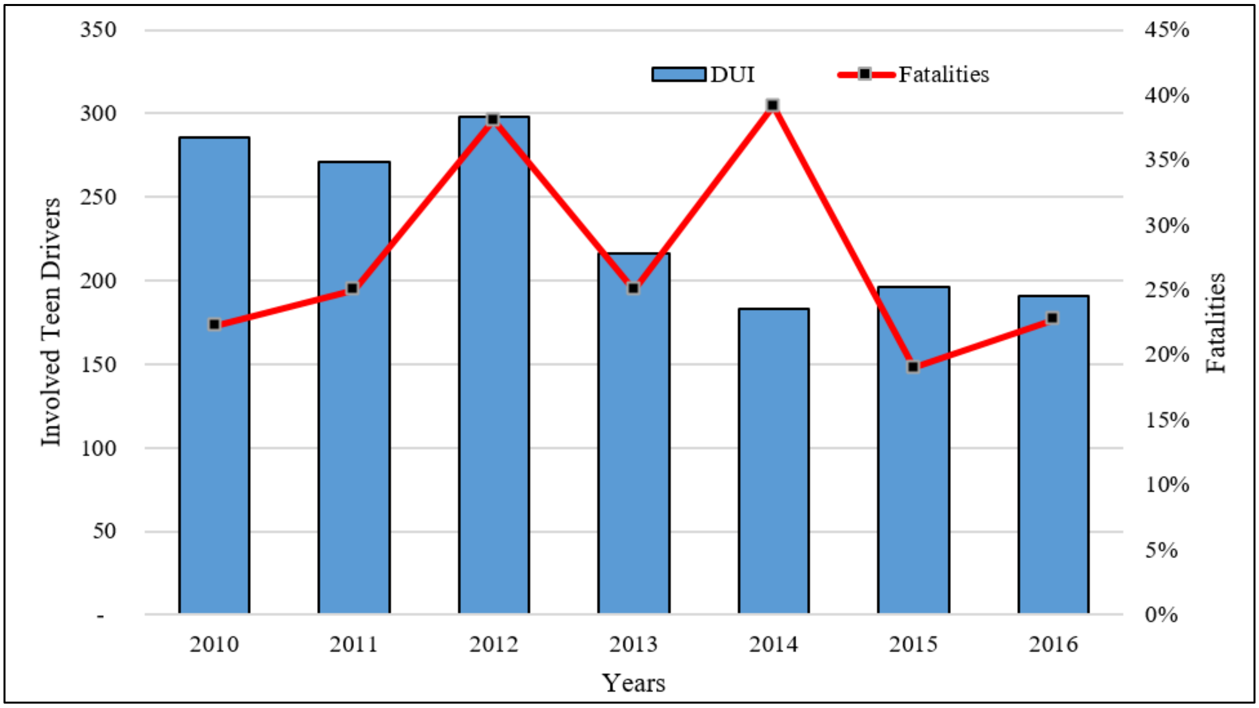


Figure 34. Teen Drivers Involved in Crashes Related to Driving Under the Influence

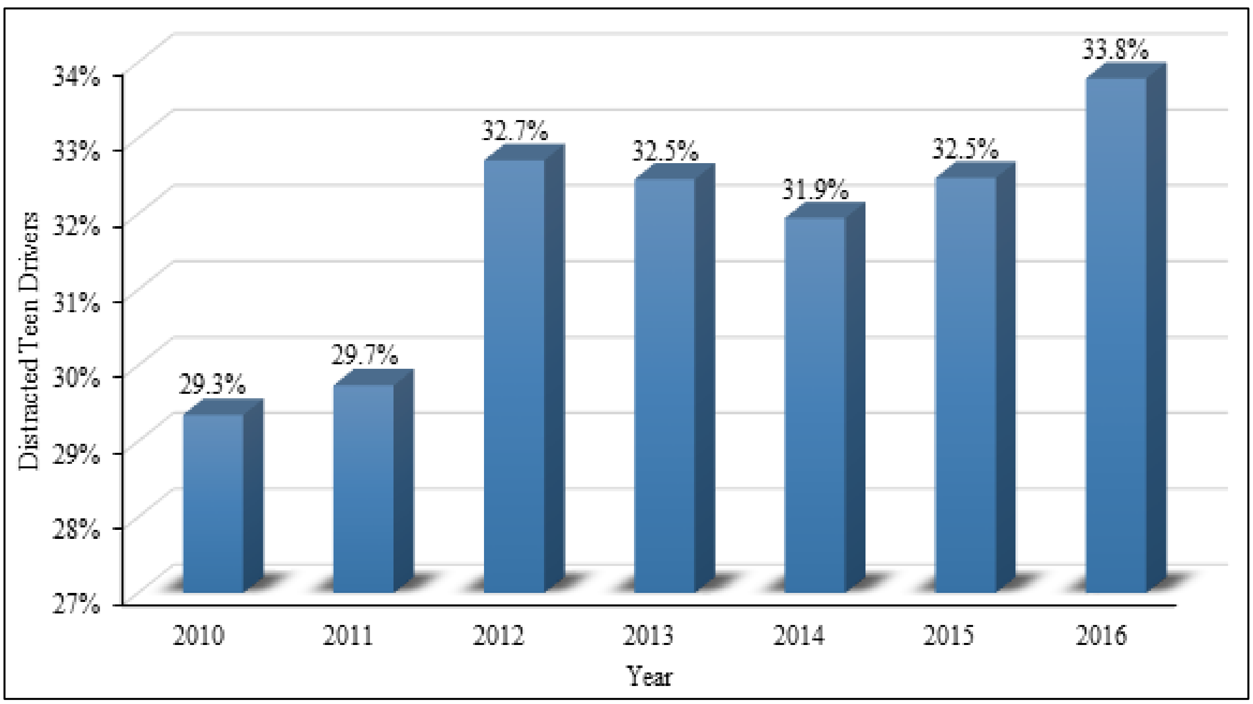


Figure 35. Percentage of Distracted Teen Drivers

Vehicle Characteristics

In this section, the vehicles driven by teen drivers involved in traffic crashes were analyzed. The vehicle characteristics included vehicle body types and age of vehicles in term of made years. The results shown in Table 18 indicate that the majority of the vehicle types involved in crashes and the type of vehicles driven by teen drivers in the six-year period were passenger cars, which represent about 65 percent of the vehicles, pickup trucks and SUVs came in the second and third level by 16.3 percent and 15.7 percent, respectively. Moreover, 60 percent and more than 58 percent of the teen driver fatalities and disabilities, respectively, occurred while they were driving passenger cars (see APPENDIX B, Table 34).

Table 18. Body Types of Vehicles Involved in Crashes That Driven by Teen Drivers

Body Type	2010	2011	2012	2013	2014	2015	2016	Total	
								No.	(%)
PC	8,504	7,823	7,468	7,408	7,263	7,251	7,673	53,390	64.69
Motorcycle, Moped, Scooter	50	67	81	55	46	62	58	419	0.51
Pickup Track	2,141	1,913	1,841	1,841	1,879	1,920	1,927	13,462	16.31
SUV	1,828	1,689	1,711	1,699	1,846	1,990	2,177	12,940	15.68
Van	327	306	285	271	269	279	260	1,997	2.42
Others[1]	36	54	45	53	54	40	39	321	0.39
Total	12,886	11,852	11,431	11,327	11,357	11,542	12,134	82,529	

In regards to the model year of the vehicle involved in crashes that were driven by teen drivers, 86 percent of those vehicles were more than five years old, as shown in Figure 36. The model years in the figure contains some controversial model years. For instance, some vehicles

[1] Other type of vehicles were driven by teen drivers included: ATV, RV, buses, farm equipment, large trucks, etc.

from 1900, 1937, 1946, 1948, and 1949 were reportedly driven by teen drivers involved in crashes. It is possible the model year numerals of these vehicles were transposed or inputted incorrectly. This perspective is probably true for other reported model year such as 1955 and 1960, or it might even go further. Since it is unknown exactly where that breakpoint is, for simplicity, they all have been retained in the analysis.

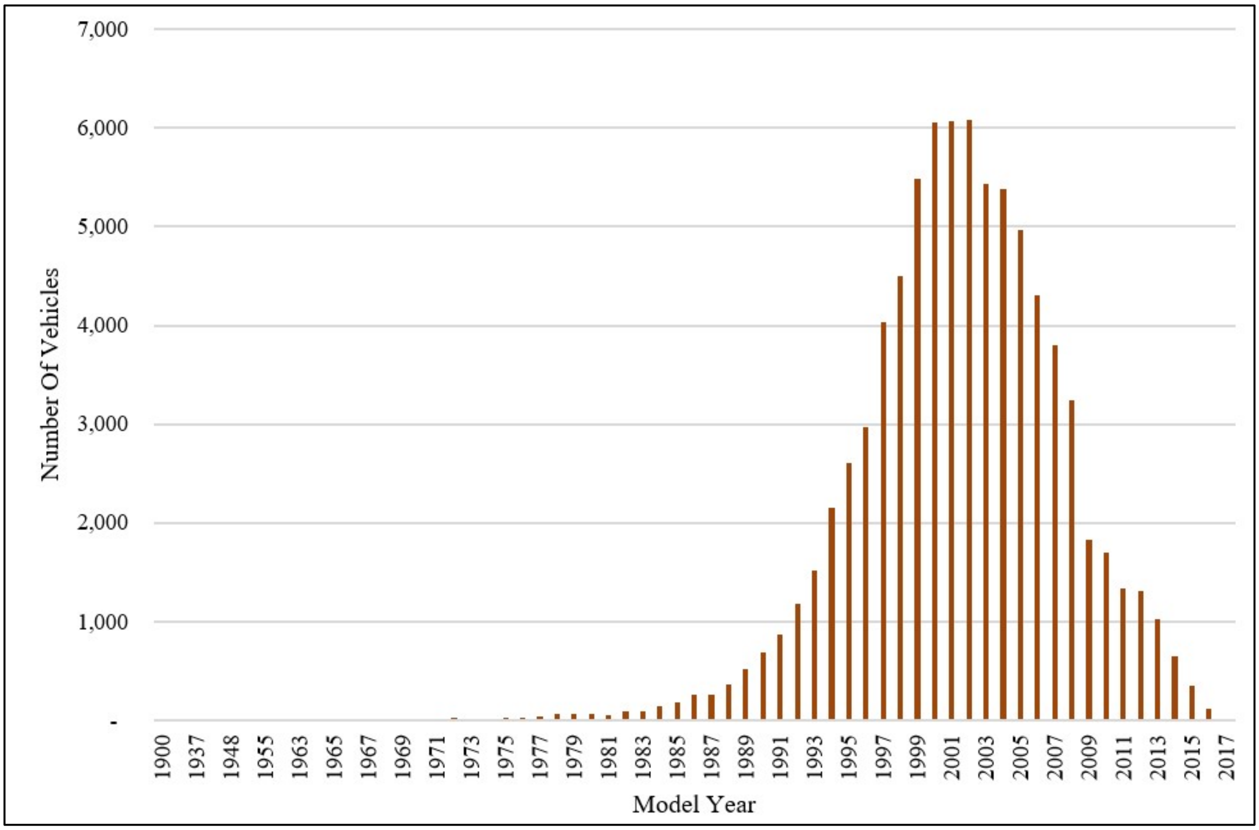


Figure 36. Model Year of Vehicles Involved in Crashes That Driven by Teen Drivers

The mode-vehicle driven by teen drivers involved in crashes was a 2002 model-year vehicle. The age of these vehicles could be as new as an eight-year-old vehicle involved in crashes that happened in 2010 or up to a 14-year old vehicle for crashes in 2016. The mean and median of vehicles driven by teen drivers involved in crashes were 2001 and 2002 model-year vehicles, respectively. One standard deviation of the model year of vehicles driven by teen drivers and involved in crashes was near ± 6 years. This is, 68.2 percent of the vehicles driven by teen drivers involved in crashes were between 1995 and 2007 model years. Further details on the results by year can be found in APPENDIX B, Table 35.

Furthermore, the percentage of passenger cars and motorcycles involved in crashes increased linearly with the age of drivers. Conversely, the percentage of larger body types decreased with driver age, as shown in Figure 37 in APPENDIX B. In terms of age and gender of teen drivers and the body type of vehicles which were driven by them, the results showed that the majority of involved pickup trucks and motorcycles were driven by male teen drivers. On the other hand, the larger number of passenger cars and SUVs involved in crashes were driven by female teen drivers. These results are shown in Table 36. The next section includes an in-depth spatial analysis of the crashes that were described in this section.

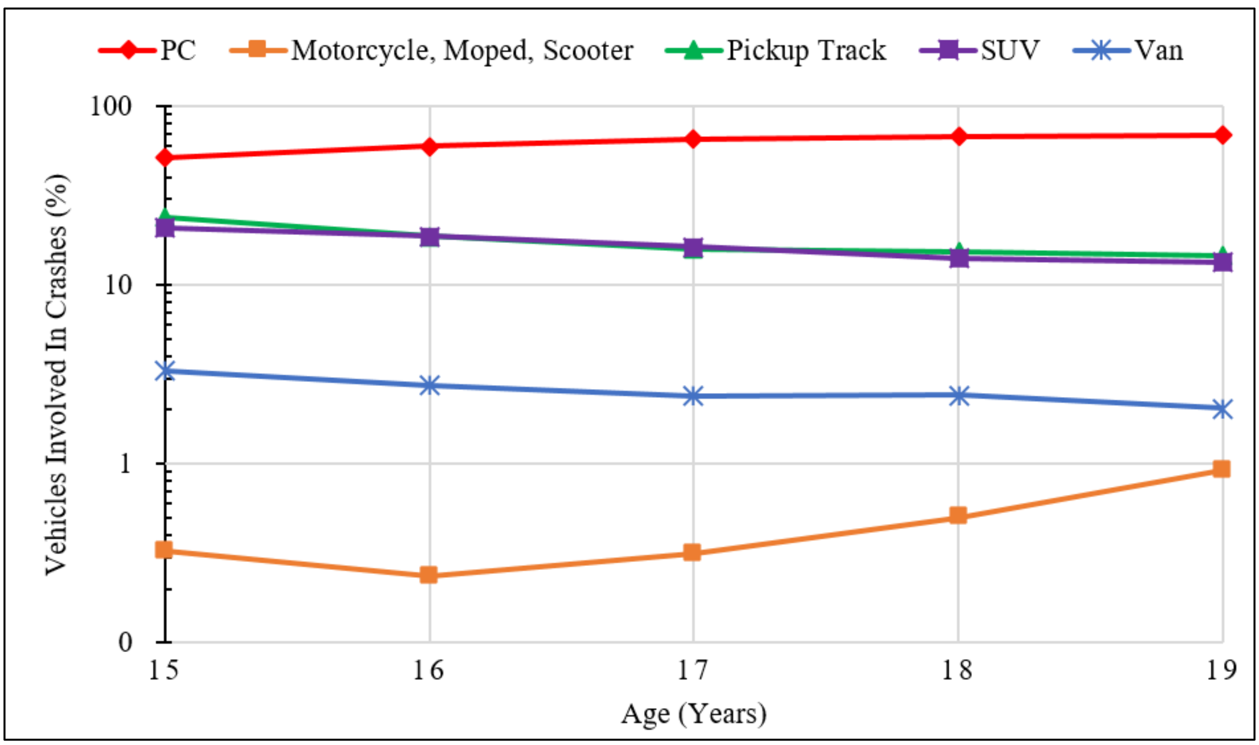


Figure 37. Type of Vehicles Involved in Crashes by Driver Age

SPATIAL ANALYSIS

The spatial analysis was conducted at two levels: statewide and at the unified school district level. These were selected as an example to conduct an in-depth investigation to show the utility of the methodology for different sized areas.

STATE LEVEL ANALYSIS

The spatial analysis at the state level was applied using the major functions in the toolsets listed in Chapter Three (Spatial Analysis section). This was due to the fact that at the state level there are ready-made subdivisions (counties), which provided a way to subdivide data and apply most of the spatial analysis functions. However, at the school district level, there were no useful subdivisions that aggregate traffic crash data, which required the use of shaped polygons to compete the analysis.

Measuring Geographic Distribution

The two primary tools that were used from the Measuring Geographic Distribution toolset to summarize the spatial distribution of crashes involving teen drivers were the Mean Center and the Directional Distribution.

Mean Center

The Mean Center is a measure of central tendency that identifies the geographic center for crash points that occurred in a defined area. To answer questions like, "Where are the fatal and non-fatal crashes involving teen drivers centered?", the mean center function was applied on the projected fatal and non-fatal crash data, as shown in Figure 38.

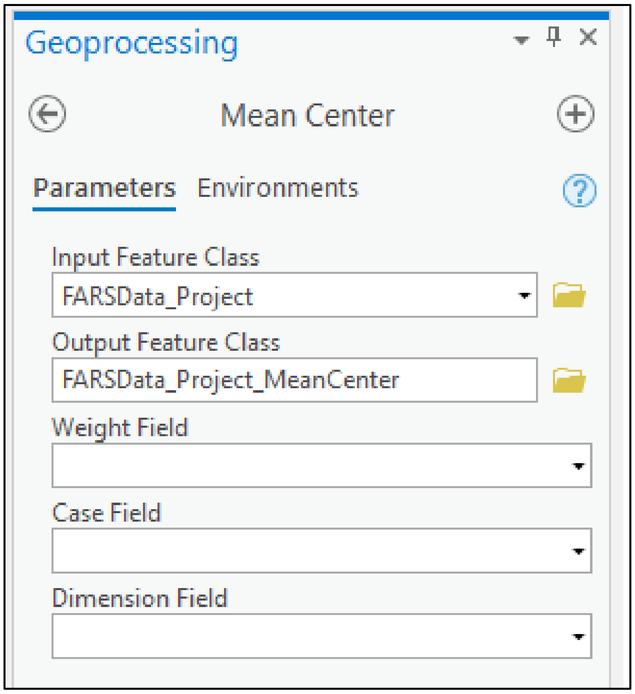


Figure 38. Mean Center Application Window

The mean center of fatal and nonfatal crashes for the study period was located in the eastern portion of the state between Sedgwick and Shawnee counties. The mean center of fatal crashes involving teen drivers was located 28.1 miles west-southwest of the mean center of non-fatal crashes, as shown in Figure 39. This result was expected since the massive number of crashes that occurred in Districts One and Five. The spatial distribution of mean centers for both fatal and non-fatal crashes by year showed that the mean centers of non-fatal crashes clustered on the northeast corner of Chase County. However, the mean centers of fatal crashes spread out among four counties McPherson, Marion, Morris, and Chase. This spatial difference was expected because the distribution of fatal and non-fatal crashes between counties was heterogeneous, as shown previously in Figure 26. More details are shown in APPENDIX C (Figure 61 and Figure 62).

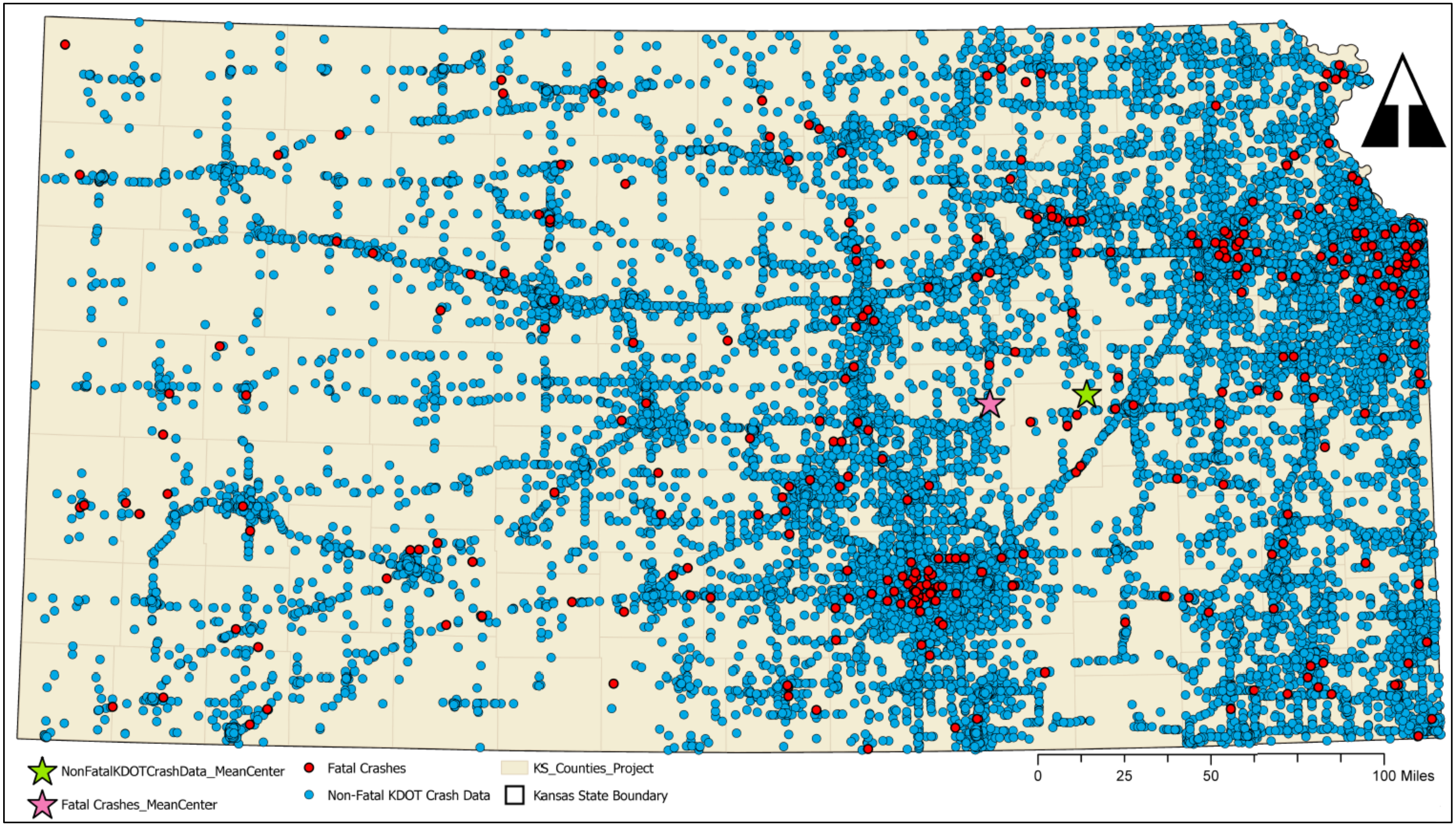


Figure 39. Mean Centers for Fatal and Non-fatal Crashes Involving Teen Drivers

Directional Distribution (or Standard Deviational Ellipse)

Measuring the spatial trends (central tendency, dispersion, and orientation) for fatal and non-fatal crash points around their means using the standard deviational ellipse encompass the orientation and directional distribution of these crashes in the study area. The standard deviational ellipse was used to compare the distributions of fatal and non-fatal crashes involving teen drivers. The results shown in Figure 40 indicate that both categories of crashes have similar trends defining a concentrated zone for the state. However, the standard distances of the standard deviational ellipse for fatal crashes (red ellipse) were greater than the standard distances of the standard deviational ellipse for non-fatal crashes (green ellipse), as shown in Table 19. This means that fatal crashes involving teen drivers were more dispersed over the study area than non-fatal crashes. Similarly, the orientation, which represents the rotation of the X-axis measured clockwise from north, for fatal crashes was greater than non-fatal crashes, but both of the ellipses give an indication of a southwest-to-northeast directional pattern.

Table 19. The Standard Distances of the Standard Deviational Ellipse for Fatal and Non-fatal Crashes

Crash Type		Fatal (mile)	Non-fatal (mile)
Standard Distance for	X-axis	284.0	260.0
	Y-axis	146.0	121.2
Rotation Angle		81.2°	73.0°

The standard deviational ellipse for both fatal and non-fatal crashes by year indicate that the means centers of non-fatal crashes were clustered while for fatal crashes were more dispersed during the study period across the state. More details are shown in Figure 63 and Figure 64 in APPENDIX C.

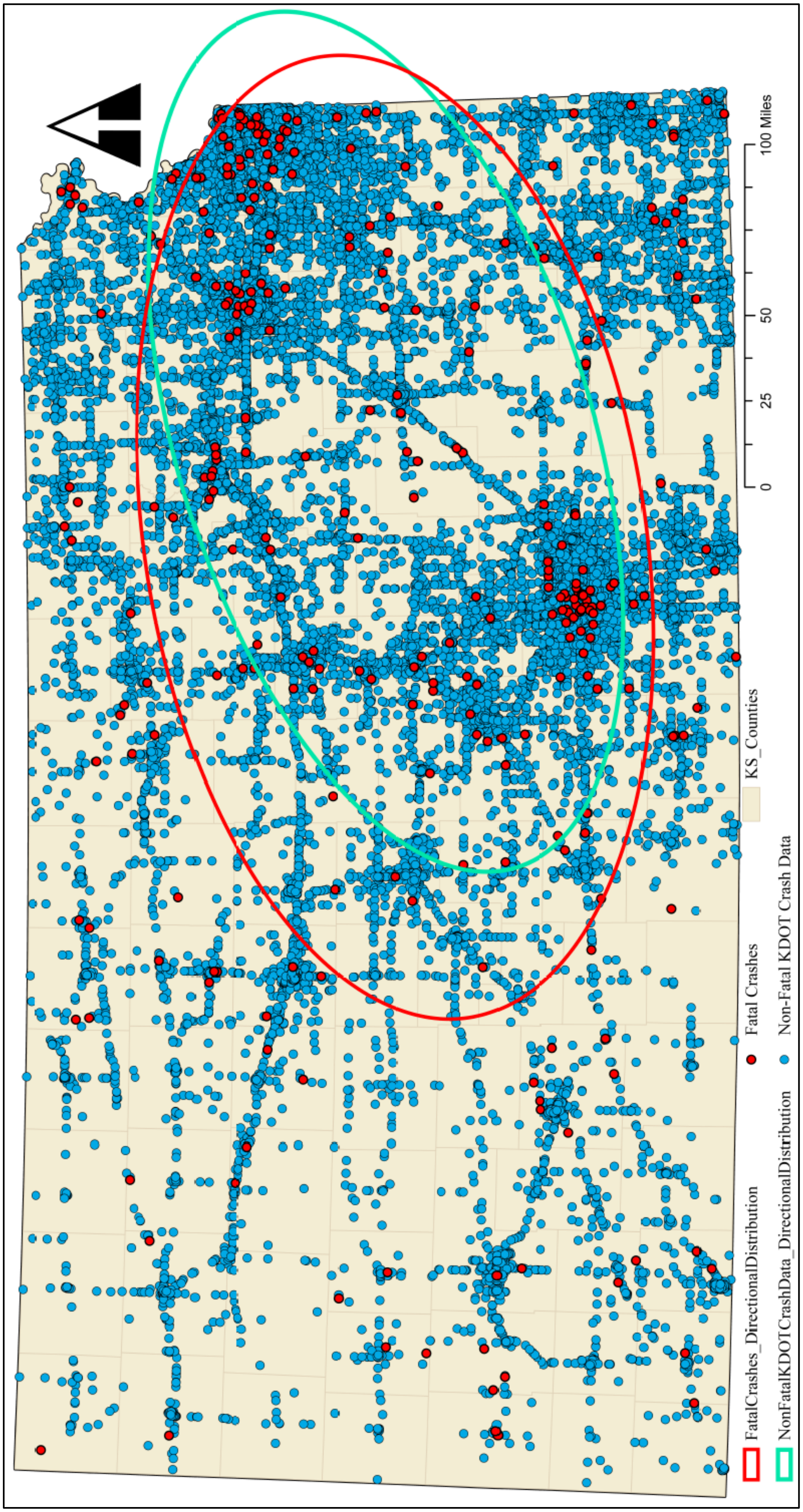


Figure 40. Directional Distribution for Fatal and Non-fatal Crashes Involving Teen Drivers

Analyzing Patterns

The Analyzing Patterns toolset comprises different tools that assist in understanding wide spatial patterns and trends of crashes involving teen drivers. The tools in this toolset are inferential statistics that use statistics to measure clustered features and facilitate comparing crash patterns for different crash types or tracking the changes in their patterns over time displayed on maps (Esri, 2014f). Using statistics to measure patterns means statistically comparing observed distributions to a hypothetical random distribution of the same number of observations over the same area (Mitchel, 2005). The tools in this toolset help to evaluate whether features or their corresponding values form a clustered, dispersed, or random spatial pattern (Pimpler, 2017). The most common tools in this toolset are the average nearest neighbor, and high/low clustering (G_i*), and spatial autocorrelation (global Moran's I).

Average Nearest Neighbor (ANN)

The ANN tool was used to compare the distribution of fatal and non-fatal crashes and also for the gender of involved drivers to find out which one is more clustered than the other. The Nearest Neighbor ratio for fatal crashes involving teen drivers was 0.678. Given the z-score of -10.873 and p-value less than 0.0001, there is a less than one–in–ten–thousand likelihood that this clustered pattern could be the result of random chance, as shown in Figure 41. The Nearest Neighbor ratio for non-fatal crashes involving teen drivers was 0.2956. Given the z-score of -362.533 and p-value less than 0.0001, indicating a less than one–in–ten–thousand likelihood that this clustered pattern could be the result of random chance. The nearest neighbor ration and z-score of non-fatal crashes are less than of the fatal crashes, which indicates that the non-fatal crashes were more clustered than the fatal crashes. Further details on the ANN results of fatal and non-fatal crashes are displayed in APPENDIX C, Figure 65 and Figure 66.

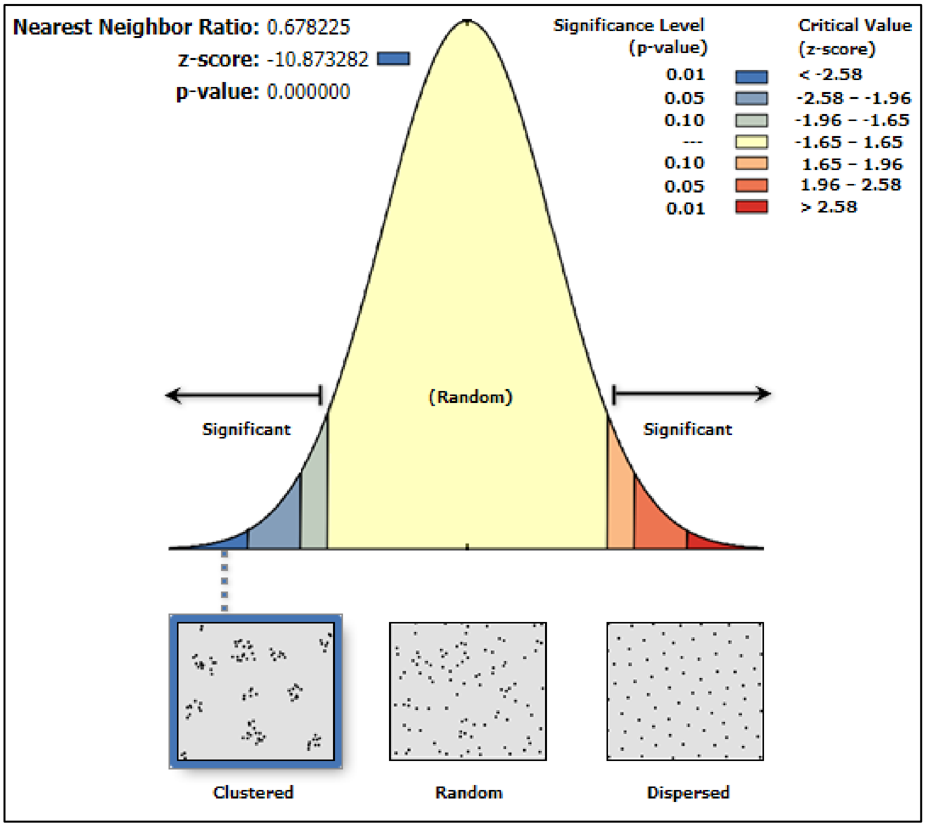


Figure 41. The Result of the Average Nearest Neighbor for Fatal Crashes

Analyzing the pattern of fatal crashes involving teen drivers based on drivers' gender was conducted. The results showed that fatal crashes for both genders were clustered but those crashes involving male teen drivers were more clustered than those involving female drivers, as shown in Table 20.

Table 20. The Average Nearest Neighbor of Fatal Crashes by Gender

Results\Gender	Male Teen Driver	Female Teen Driver
Nearest Neighbor Ratio	0.695	0.837
z-score	-8.525	-3.109
p-value	0.0000	0.0019

High/Low Clustering (Getis-Ord General G)

The General G can tell statistically whether clustering for either high values or low values exists in the study area, but it does not identify the location of the cluster. The analysis was conducted based on counties. Therefore, the contiguity (Contiguity_Edges_Corners) option in the conceptualization field was selected (see Figure 67 and Figure 68 in APPENDIX C for the analysis report). For the standardization field, the row standardization parameter was selected, as shown in Figure 67 and Figure 68 in APPENDIX C.

The Getis-Ord General G analysis outcomes for fatal crashes (shown in Figure 42), and for non-fatal crashes involving teen drivers were found to have z-scores statistically significant at the 99 percent level of confidence because they were larger than 2.58 and the p-value was less than 0.01. When the z-score value is positive, it means that the observed General G value is larger than the expected General G value. Positive z-scores greater than 1.65 is an indication of clustering high values in the study area. But when the z-score value is less than -1.65, it is an indication for the low values clustering in the study area. That means for this case, high values of fatal and non-fatal crashes were clustered in the state. This conveys that fatal and non-fatal crashes involving teen drivers were clustered in counties like Johnson and Sedgwick, for example, which had the highest number of those crashes (high values).

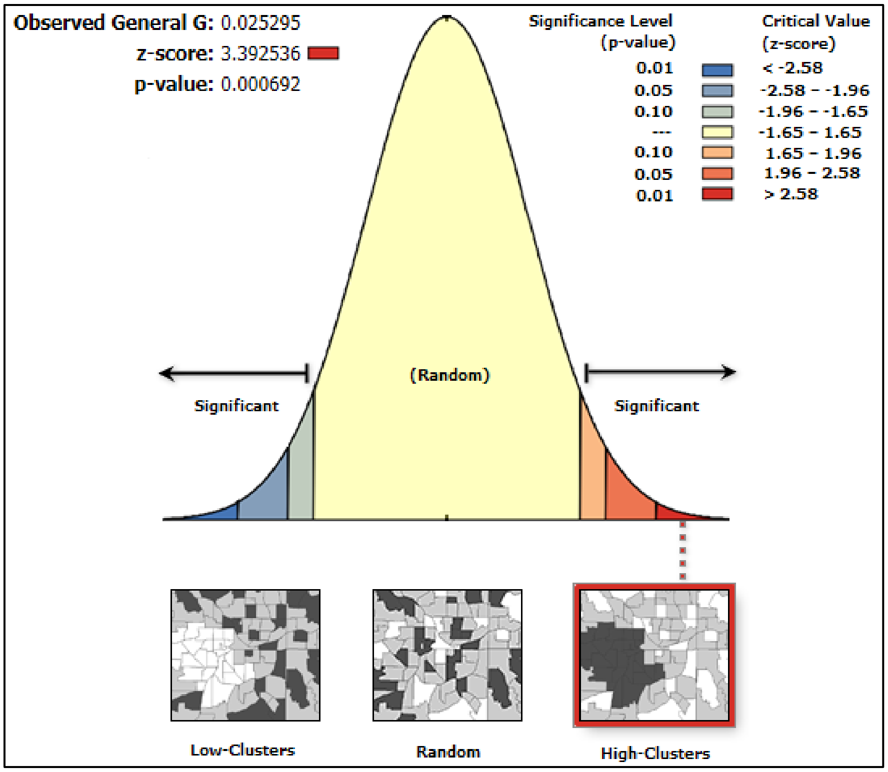


Figure 42. The High/Low Clustering (Getis-Ord General G) Results on Fatal Crashes

The null hypothesis for the High/Low Clustering (General G) statistic states that feature values are not spatially clustered, which implies that the crashes were randomly distributed among the counties. But the resultant p-values were small and statistically significant. Therefore, the null hypothesis was rejected. The high-value cluster was expected because the crash distribution across the study area was not evenly distributed and most of the crashes were. Further details on the statistical analysis results are shown in Table 21.

Table 21. The High/Low Clustering (General G) for Fatal and Non-fatal Crashes

Results\Crash Severity	Fatal Crashes	Non-fatal Crashes
Observed General G Values	0.0153	0.0253
Expected General G Value	0.0096	0.0096
z-score	3.8020	3.3925
p-value	0.000144	0.000692

Spatial Autocorrelation (Global Moran's I)

Global Moran's I measures spatial autocorrelation based on feature locations and attribute values to evaluate whether the data are clustered, dispersed, or random. The Global Moran's I analysis outcomes for fatal crashes (see Figure 43), and non-fatal crashes (see Figure 69 in APPENDIX C) involving teen drivers display that the z-scores were statistically significant at the 99 percent level of confidence because they are larger than 2.58 and p-value was less than 0.01. When the z-score value is larger than zero, it means that the observed Global Moran's I value is larger than the expected Global Moran's I value, which indicates feature values are clustered in the study area.

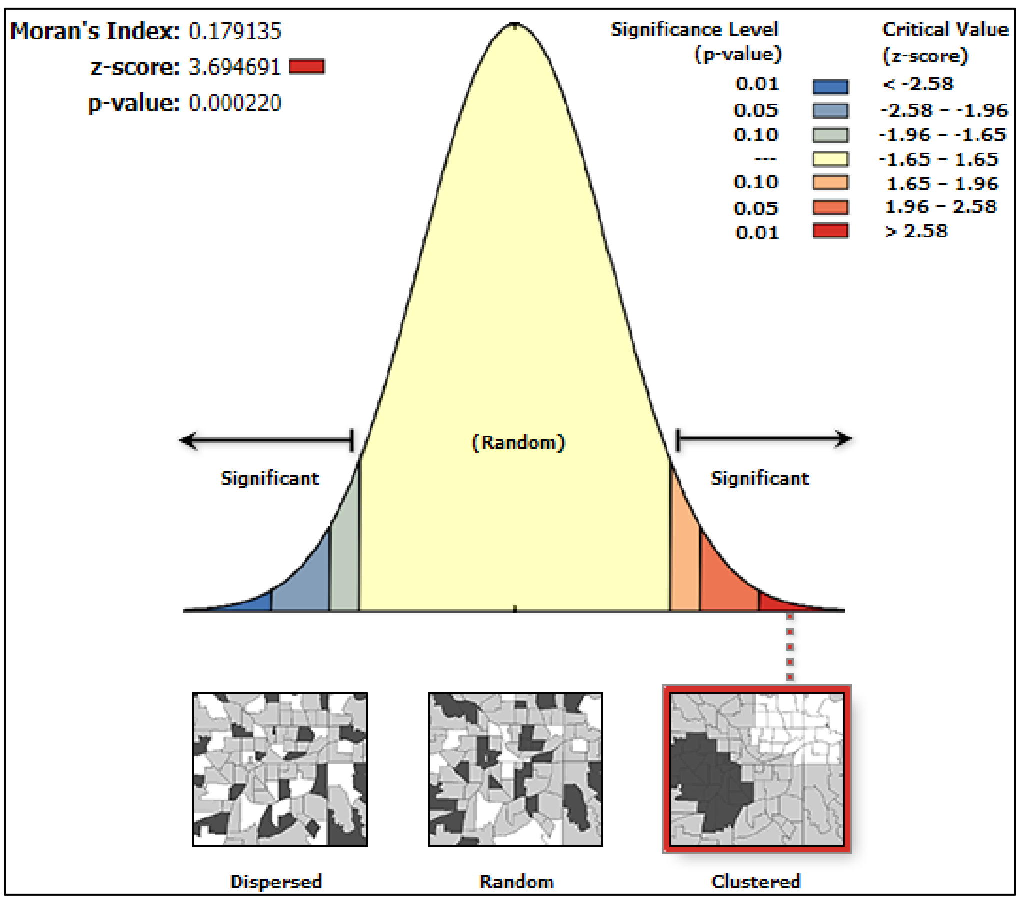


Figure 43. Spatial Autocorrelation Results on Fatal Crashes

The null hypothesis for the Spatial Autocorrelation (Global Moran's I) statistic states that feature values are not spatially clustered, which implies that the crashes are randomly distributed among the counties. But the resultant p-values were small and statistically significant in the level of confidence 99 percent, as shown in Table 22. Therefore, the null hypothesis was rejected, and this gave a signal to further investigate the location and contributing factors behind the clustering pattern. The table also shows that non-fatal crashes involving teen drivers were more clustered than fatal crashes.

Table 22. The Global Moran's I for Fatal and Non-fatal Crashes

Results\Crash Severity	Observed Global Moran's I	Expected Global Moran's I	z-score	p-value
Fatal Crashes	0.17914	-0.0096	3.69469	0.00022
Non-fatal Crashes	0.14746	-0.0096	3.2727	0.00107

Mapping Clusters

The tools in the previous section (Analysis Pattern) were answering questions such as "are there statistically significant clustering or dispersion patterns in the study area?" In this section, the mapping clusters tools step beyond answering "Yes" or "No" to this question. Instead, they answer questions such as "where is the statistically significant clustering?" or "where are the statistically significant hotspots and coldspots in the study area?" and visualize the patterns of the targeted datasets by creating a feature class in the map area.

The most commonly used and well-known tools in this toolset are: Optimized Hot Spot Analysis, and Optimized Cluster and Outlier Analysis.

Optimized Hot Spot Analysis

This tool identifies clusters of high values (hotspots) and clusters of low values (coldspots) for a set of weighted features within a specified distance. The visualized results of the Optimized Hotspot Analysis for fatal crashes involving teen drivers are shown in Figure 44. The map shows two hotspots with different levels of significance. One of the hotspots included Sedgwick County and its neighbor counties, and the other hotspot was 11 counties in the northeast corner of the state, which comprise the most populous counties in the northeast part of the state (Douglas, Johnson, Leavenworth, Shawnee, and Wyandotte) and some counties around them. However, no coldspots had been identified, and the rest of the state was classified as "Not Significant," which means the fatal crashes that occurred in those locations distributed randomly and there were no significant patterns.

This analysis validated all the 312 fatal crashes, and integrated them in 907 weighted hexagonal polygons. The size of created hexagon polygons was about 12.3 miles by 10.7 miles. The optimal fixed threshold distance based on peak clustering was found at 112,607 feet. In this threshold distance, only 0.2 percent of the features had less than eight neighbors, which is a good indication of the accuracy of the results. Further details on the report could be found in APPENDIX C, Figure 70.

For non-fatal crashes involving teen drivers, the results of the Optimized Hotspot Analysis is shown in Figure 45. The results display another hotspots beside the hotspots at most of the locations identified for fatal crashes. One hotspot is located in three counties (Geary, Pottawatomie, and Riley), which include the city of Manhattan and the Fort Riley Army Base. Another hotspot was in the Sedgwick County and skewed east to encompass Butler County.

.

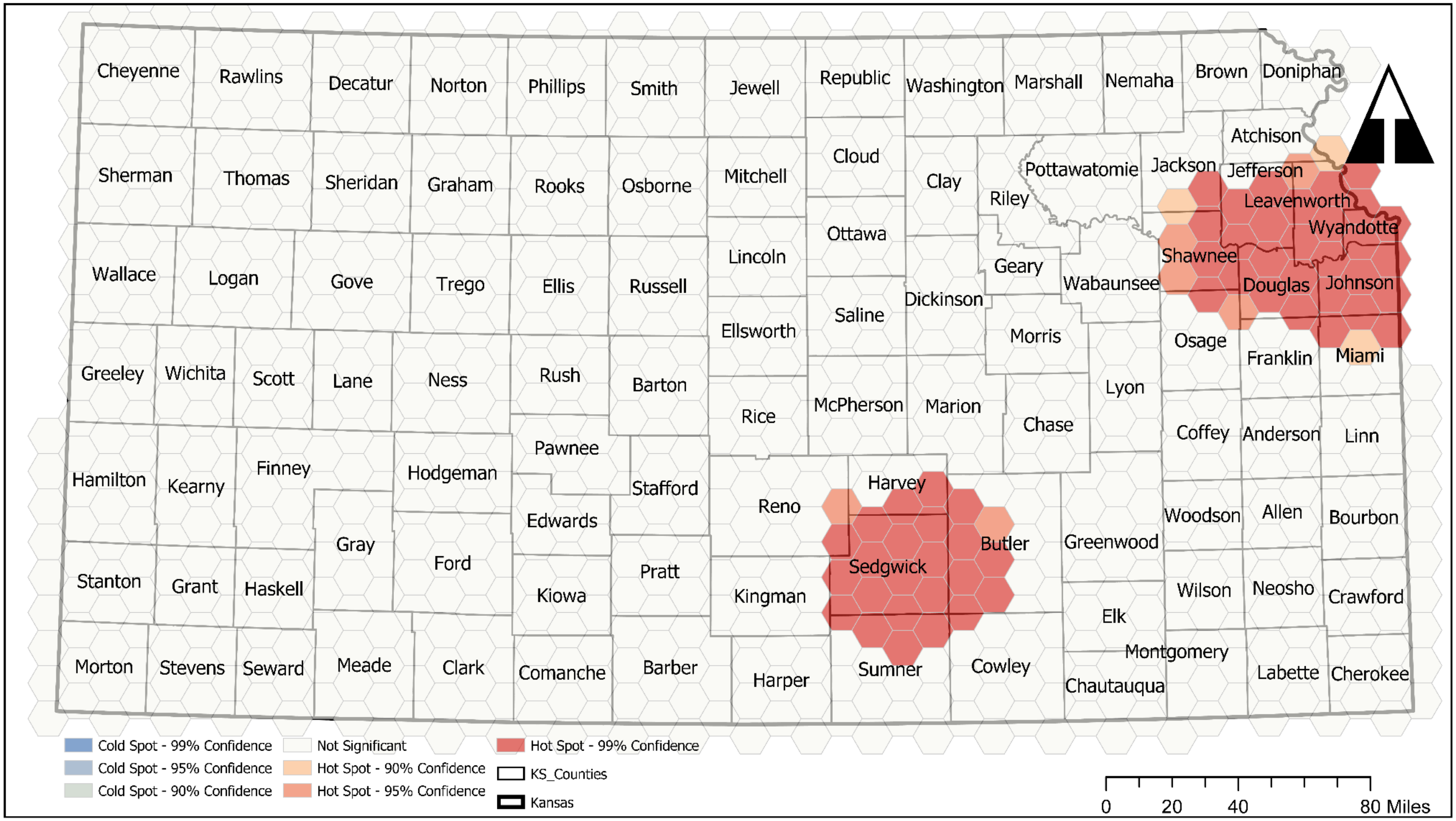


Figure 44. The Optimized Hot Spot Analysis Result for Fatal Crashes Invovling Teen Drivers

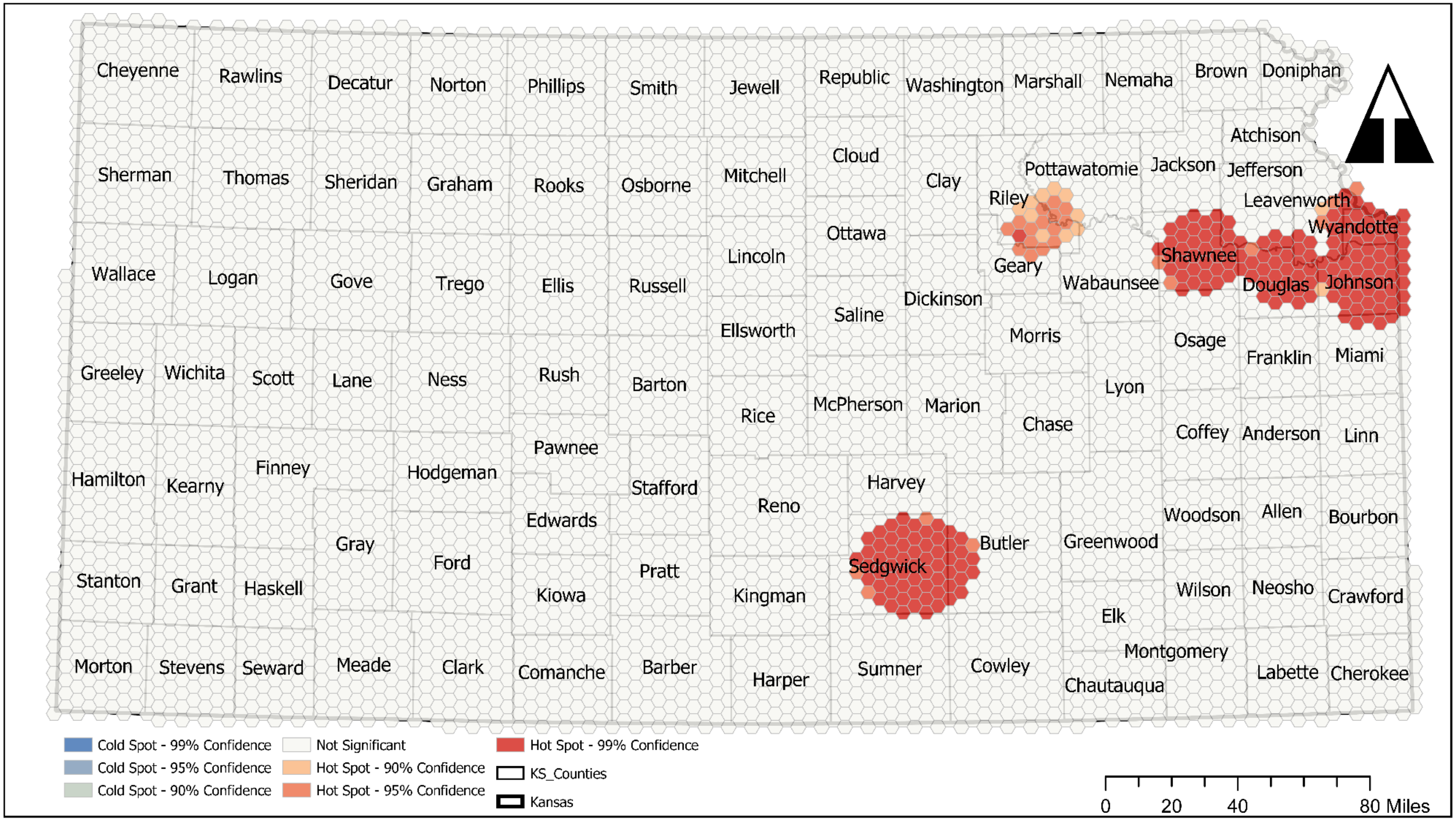


Figure 45. The Optimized Hot Spot Analysis Result for Non-Fatal Crashes Invovling Teen Driver

The other hotspot group was series of adjacent hotspots concentrated in Douglas, Johnson, Shawnee, and Wyandotte counties in the northeast portion of the state.

This analysis validated all the 72,591 non-fatal crashes, and it integrated them in 5,943 weighted hexagonal polygons. The size of the created hexagon polygons was about 4.7 miles by 4.1 miles. The optimal threshold distance based on peak clustering was found at 64,368 feet. In this threshold distance, all the features had at least eight neighbors, which is an excellent indication for the accuracy of the results. Further details on the report could be found in APPENDIX C, Figure 71

Optimized Cluster and Outlier Analysis (Anselin Local Moran's I)

This tool is the local version of Moran's I, which investigates what is happening in features that are directly adjacent to the targeted feature. In other words, it examines the probability that the similarity that existed between a feature, and its neighbors did not occur randomly. It identifies clusters of high or low values (crashes in this case) and spatial outliers of a set of weighted features (counties in this case). A high positive value of Local Moran's I creates a high positive z-score, which indicates that the targeted county is bordered by similar values (cluster), which could be either high values or low values. On the contrary, a very negative value of Local Moran's creates a very negative z-score, which indicates that the targeted county is bordered by dissimilar values (outlier), which they could be either High-Low or Low-High outliers.

The cluster and outlier analysis results for fatal crashes involving teen drivers are shown in Figure 46. There were 39 outlier features that they were statistically significant. The three red features fall in the High-Low category of outliers. Two of these features were located in Cheyenne and Sherman counties and the other one was located between Pawnee and Edwards counties.

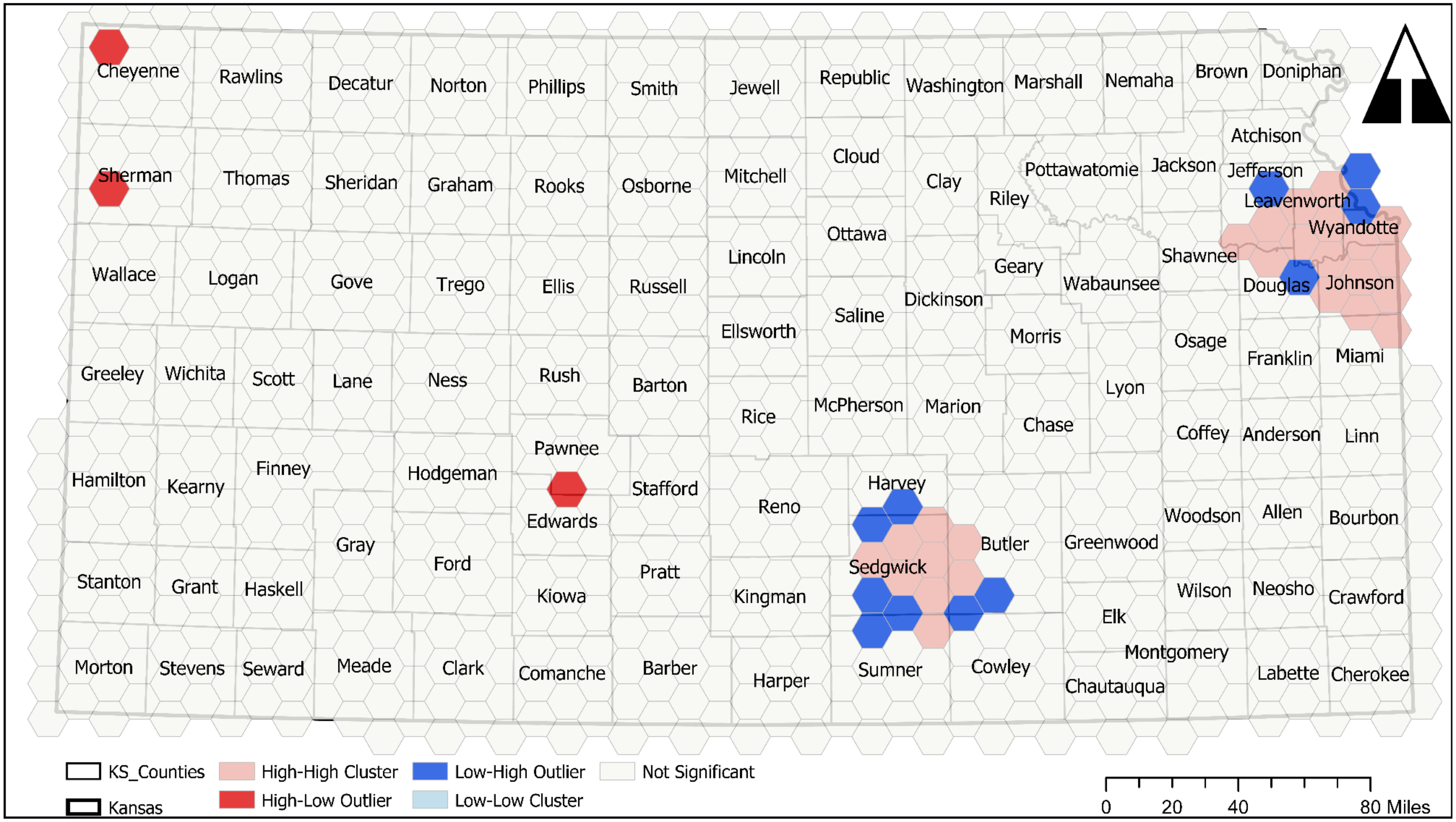


Figure 46. The Optimized Outlier Analysis Result for Fatal Crashes Involving Teen Driver

In these locations, fatal crashes involving teen drivers were high but the surrounding counties had a low or not statistically significant number of fatal crashes. This means these locations had a unexpectedly high number of crashes compared to their neighboring features. The other outlier category is Low-High outliers, which are shown in dark blue polygons. These 11 features had a low number of fatal crashes but their surrounding counties had a high or not statistically significant number of fatal crashes. This means that these locations had unexpectedly fewer fatal crashes involving teen drivers in comparison to their neighborhoods.

In term of clusters, 25 features were identified as High-High clustered features and no features were identified as Low-Low clustered features. The High-High clusters mean that these features had a high number of fatal crashes involving teen drivers and their surrounding features also had a statistically significant high number of fatal crashes. The analysis report showed that all the 907 weighted hexagon polygons were used. The optimal fixed threshold distance based on peak clustering was found at 112,607 feet. In this threshold distance, only 0.6 percent of the features had less than eight neighbors, which is a good indication of the accuracy of the results. Further details on the report could be found in APPENDIX C, Figure 72.

The cluster and outlier analysis results for non-fatal crashes involving teen drivers are shown in Figure 47. There were 26 High-Low outliers shown in bright red in non-fatal crashes, and they were mostly in the western part of the state. This means 26 features had an unexpectedly high number of non-fatal crashes compared to their neighboring features. However, 52 features were identified as Low-High outliers, which are shown in dark blue hexagonal polygons located frequently around hotspots.These features had a low number of non-fatal crashes, but their surrounding features had a high or not statistically significant number of non-fatal crashes.

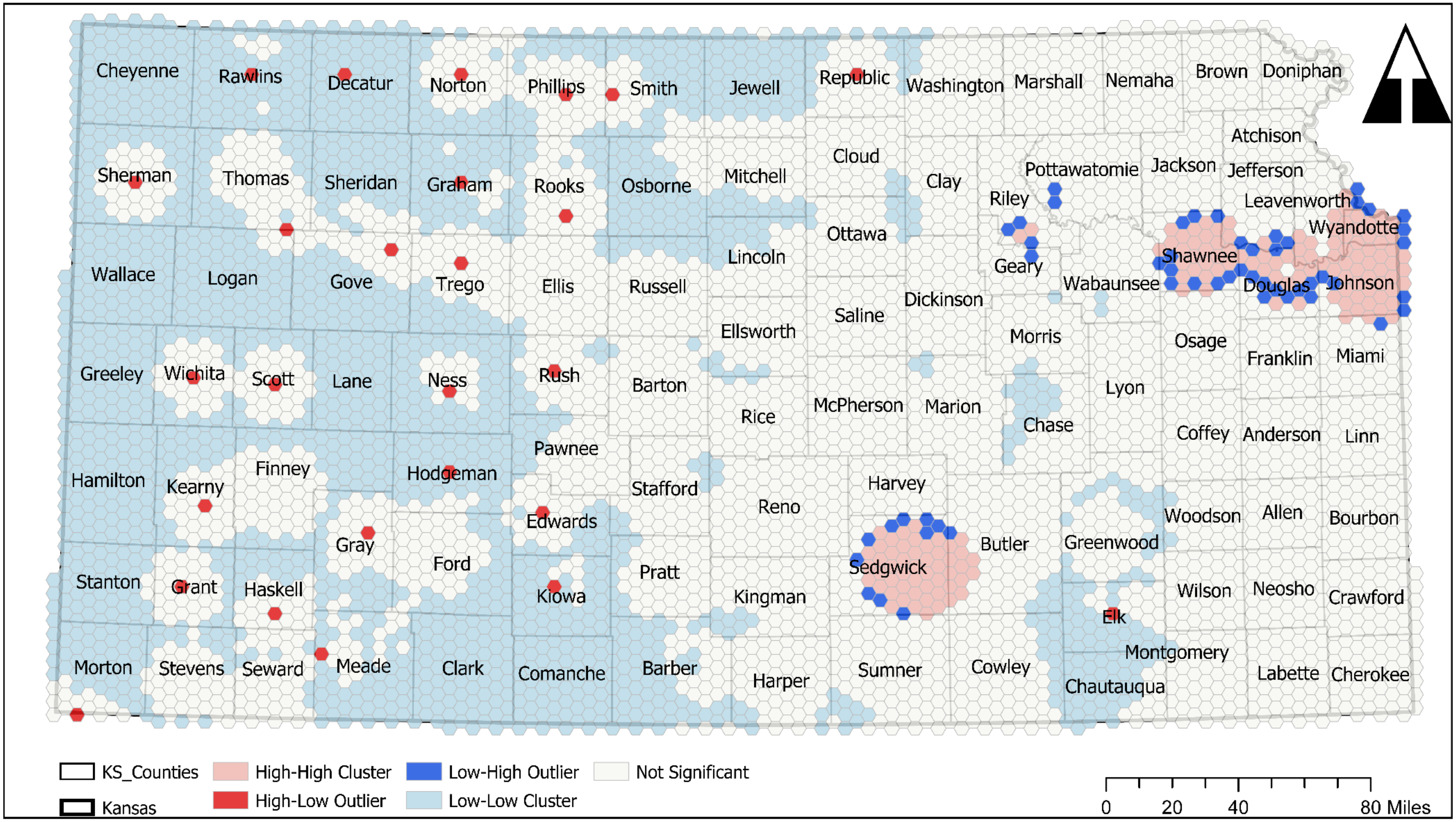


Figure 47. The Optimized Outlier Analysis Result for Non-Fatal Crashes Involving Teen Drive

This means that these features unexpectedly had fewer non-fatal crashes involving teen drivers in comparison to their neighbors.

In term of clusters, 2,047 features were identified as Low-Low clustered features or coldspots (most of them located in the western part of the state). The Optimized Hot Spot Analysis tool did not identify any coldspot features. Furthermore, 149 features were identified as High-High clustered locations or hotspots, which were almost in the same locations identified by the Optimized Hot Spot Analysis tool. The Low-Low clusters mean that these locations had a low number of non-fatal crashes involving teen drivers and their surrounding locations also had a statistically significant low number of non-fatal crashes, and vice versa for the High-High clusters. The analysis report showed that all the 5,943 weighted hexagonal polygons were used. The optimal threshold distance was found at 64,368 feet. At this threshold distance, all of the features had at least eight neighbors, which is a robust indicator for the accuracy of the results. Further details on the report can be found in APPENDIX C, Figure 73.

Beyond analyzing the distribution, pattern, and clustering of geographic features, ArcGIS can be used to identify and measure the relationships between features using modeling spatial relationships, which helps to answer questions such as: "Why do crashes involving teen drivers occur where they do?" "Where are these crashes more likely to occur in the future?"

Modeling Spatial Relationships

The regression analysis tools in ArcGIS to model spatial relationships are Ordinary Least Squares and Geographically Weighted Regression.

Ordinary Least Squares (OLS)

This global model was used to create a single equation that describes the relationship between a dependent variable (the number of traffic crashes involving teen drivers from 2010 to 2016) and each of the explanatory variables. There were 18 exploratory variables that were examined by the exploratory regression in order to select appropriate variables for the OLS model. Table 23 shows the first outcome of the exploratory regression, which includes the threshold criteria and also the number of trials and number and percentage of time that the trials passed the threshold criteria (or criterion cutoff).

Table 23. Percentage of Search Criteria Passed

Search Criterion Cutoff	Trials	#Passed	%Passed
Min Adjusted R-Squared > 0.50	11,706	11,684	99.81
Max Coefficient p-value < 0.05	11,706	323	2.76
Max VIF Value < 7.50	11,706	1,865	15.93
Min Jarque-Bera p-value > 0.10	11,706	64	0.55
Min Spatial Autocorrelation p-value > 0.10	28	24	85.71

The proposed OLS models by the exploratory regression tool are shown in APPENDIX D, Figure 81. These models were listed based on the number of exploratory variables and then the models that had the highest adjusted R-squared results. However, not all the listed models were satisfied with all the threshold criteria. Therefore, investigating the significance of each exploratory variable is the next step to select proper variables for more an in-depth investigation.

The significance of the exploratory variables shown in Table 24 defines how statistically significant each variable was during analyzing every possible combination in the Significant (%)

column and how stable variable relationships were by examining the Negative (%) and Positive (%) columns. The Strong candidate variables are those variables that were significant over 50 percent of the time (Esri Events, 2018). Accordingly, the first six variables in Table 24 were selected, and they are listed below:

- The average number of passenger cars (LNLOAVG_PC);
- Miles of rural non-state roads in a county;
- The population of teens (LNLOT_POP);
- The population of counties (LNPOP);
- Number of High Schools (HIGHSCHOOL); and
- Average DVMT on all types of roads (DVMT_ALL).

However, the only model (see APPENDIX D, Figure 81) that includes these variable and satisfies the VIF, Jarque-Bera p-value, and adjusted R-squared threshold criteria is the model that contains:

- Miles of rural non-state in a county (NONSTATE_RD)[1]; and
- The average number of passenger cars (LNLOAVG_PC).

Therefore, these two explanatory variables are the only variables that qualified to be in the reduced OLS and GWR models.

---

[1] The data source of the miles of rural non-state in a county (NONSTATE_RD), and the average number of passenger cars (LNLOAVG_PC) was provided by KDOT.

Table 24. Summary of Variable Significance from the Exploratory Regression

Variable	Significant (%)	Negative (%)	Positive (%)
LNLOAVG_PC	100.00	0.00	100.00
NONSTATE_RD	94.26	0.49	99.51
LNLOT_POP	91.45	0.00	100.00
LNPOP	74.49	23.95	76.05
HIGHSCHOOL	52.04	100	0.00
DVMT_ALL	51.55	0.00	100.00
AVG_TRUCK	38.89	24.67	75.33
DVMT_NONSTATE	30.96	22.19	77.81
NLABOR_OVER15	30.24	43.8	56.20
LNCOMMUT_WORK	29.78	37.83	62.17
NMALE_OVER15	25.86	66.50	33.50
POP18_24NO_HIGHSCH	24.24	77.91	22.09
ALL_ROAD	23.26	26.62	73.38
NFEMALE_OVER15	22.63	40.16	59.84
AVG_PRECIPT	15.14	3.23	96.77
AV_HOUSEH_INCOM	11.19	96.44	3.56
POST_SECNDRY	9.72	59.54	40.46
UNDER_POV_LEV	1.31	44.11	55.89

OLS was applied using the two exploratory variables that passed most of the significant threshold criteria of the exploratory regression. Figure 48 shows the LNLOCRASH (the number of traffic crashes involving teen drivers) variable stated as the dependent variable and miles of rural non-state in a county (NONSTATE_RD), and the average number of passenger cars (LNLOAVG_PC) as explanatory variables.

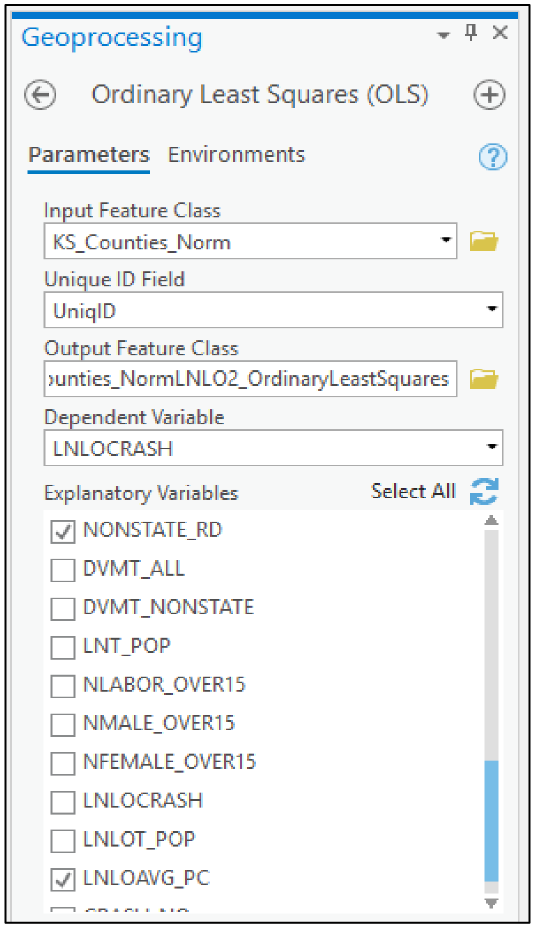


Figure 48. The OLS Application Window

The statistical report (see Table 25) shows both the Multiple R-Squared and Adjusted R-Squared values were higher than 90 percent, and this was a significant reflection of the model performance. The Adjusted R-Squared value of 0.91 indicates that the model explains approximately 91 percent of the variation in the dependent variable. The resultant model is shown in Table 26. The most critical parameters in the table are Coefficient, Probability (p-value), and VIF. Both of the coefficients have a positive relationship with the dependent variable, which is the number of crashes involving teen drivers. Thus, the more rural non-state road miles and the higher number of passenger cars, the more crashes involving teen drivers are expected. The p-value shows that the exploratory variables are statistically significant for the model.

Table 25. The Statistical Report of the OLS Regression

Multiple R-Squared	0.9101	Adjusted R-Squared	0.9084
Joint F-Statistic	516.4001	Prob.(>F), (2,102) dof	<0.0001*
Joint Wald Statistic	774.8932	Prob.(>chi-squared), (2) dof	<0.0001*
Koenker (BP) Statistic	8.6422	Prob.(>chi-squared), (2) dof	0.0133[1]
Jarque-Bera Statistic	178.9214	Prob.(>chi-squared), (2) dof	<0.0001*

Table 26. The Resultant Model from the OLS Regression

Variable	Intercept	NONSTATE_RD	LNLOAVG_PC
Coefficient	-1.065522	0.019208	1.805782
Std. Error	0.056927	0.007131	0.060281
Probability	<0.00001*	0.008261*	<0.00001*
Robust SE	0.078023	0.007178	0.065401
Robust Pr.	<0.00001*	0.008683*	<0.00001*
VIF	--------	1.086884	1.086884

The resultant model from the OLS regression could take the form below:

$$LNLOCRASH = -1.0655 + 0.0192\,(NONSTATE\_RD) + 1.8058\,(LNLOAVG\_PC)$$

For transforming back to directly receive the expected number of crashes, the model takes the form below:

$$E(y) = e^{10^{\left(-1.0655+0.0192\left(\frac{\sqrt{x_1}}{\ln x_1}\right)+1.8058\,(\log(\ln x_2))\right)}}$$

[1] An asterisk next to a number indicates a statistically significant p-value (p < 0.01)

Where $E(y)$ is the expected number of crashes, x_1 is miles of non-state roads, and x_2 is the number of passenger cars.

In Table 25, the Koenker (BP) Statistic is statistically significant, and similarly, the robust probability (see Table 26) is statistically significant. Therefore, the null hypothesis is rejected and there is a nonstationary condition in the model, which is expected as mention before. This is, the relationships between the number of crashes involving teen drivers and exploratory variables change across the study area. One or both of the exploratory variables might be a significant predictor of the number of crashes involving teen drivers in some counties, but perhaps weak predictor in other counties.

The VIF were less than 7.5, which means the variables were inconsistent in predicting the number of crashes. The Joint F-Statistic and Joint Wald Statistic p-values (see Table 25) supported that the model was statistically significant. The OLS residuals were mapped and shown in Figure 49, indicated the over predictions in blue and under predictions in red. Since the Jarque-Bera Statistic's p-value was statistically significant, the null hypothesis was rejected. This means there was heteroscedasticity because of influential outliers in the data, as shown in the map and scatterplot in Figure 49 and Figure 50, respectively. The red colored county on the map is the red dot on the scatter plot, which represents Chase County. This indicates that the model under predicting the number of crashes involving teen drivers in Chase County and the actual number were larger than the model predicted. However, the blue colored counties represent the counties where the model is over predicting the number of crashes, which means in these counties the actual number were smaller than the model predicted.

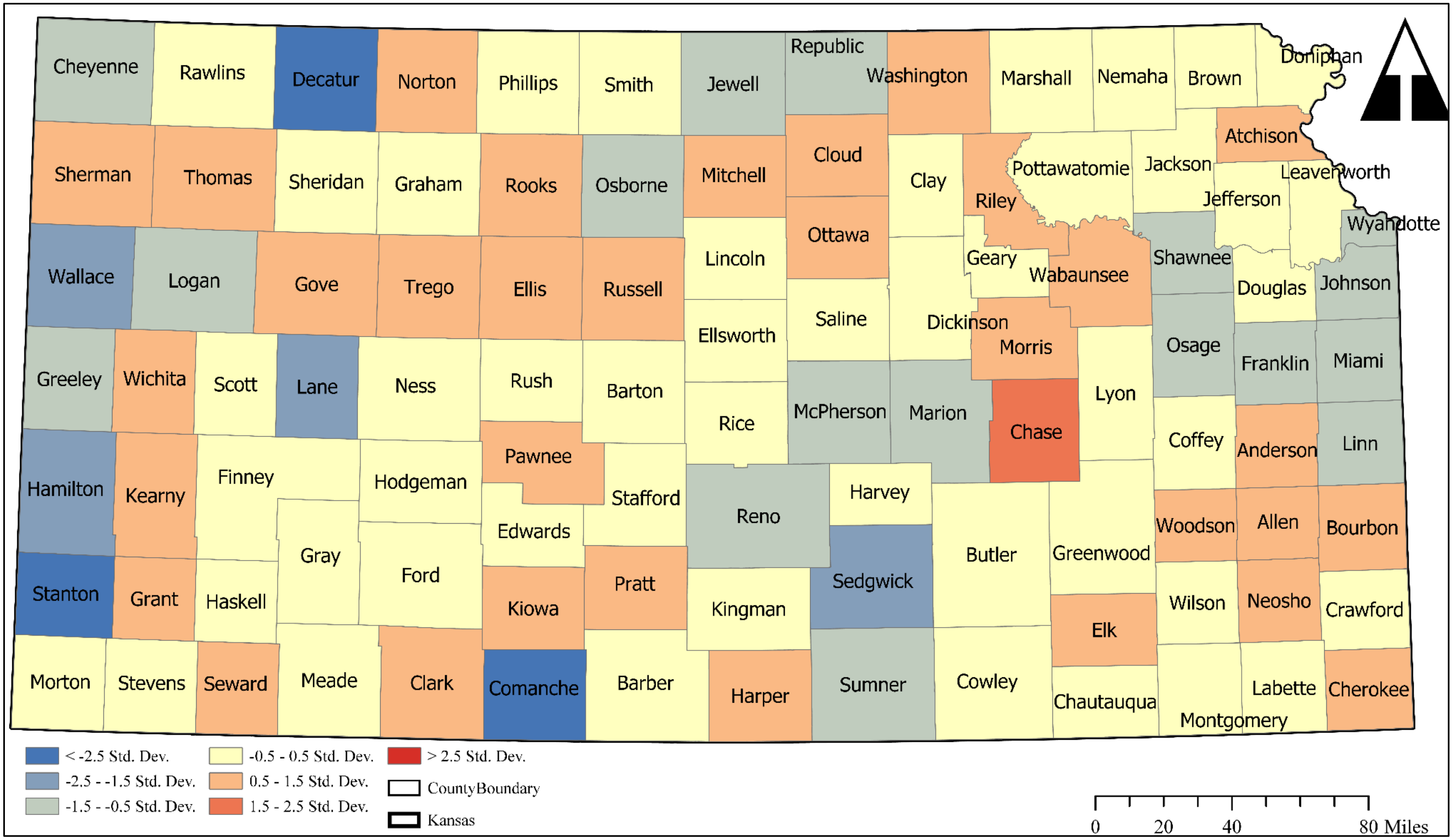


Figure 49. The OLS Mapped Residuals

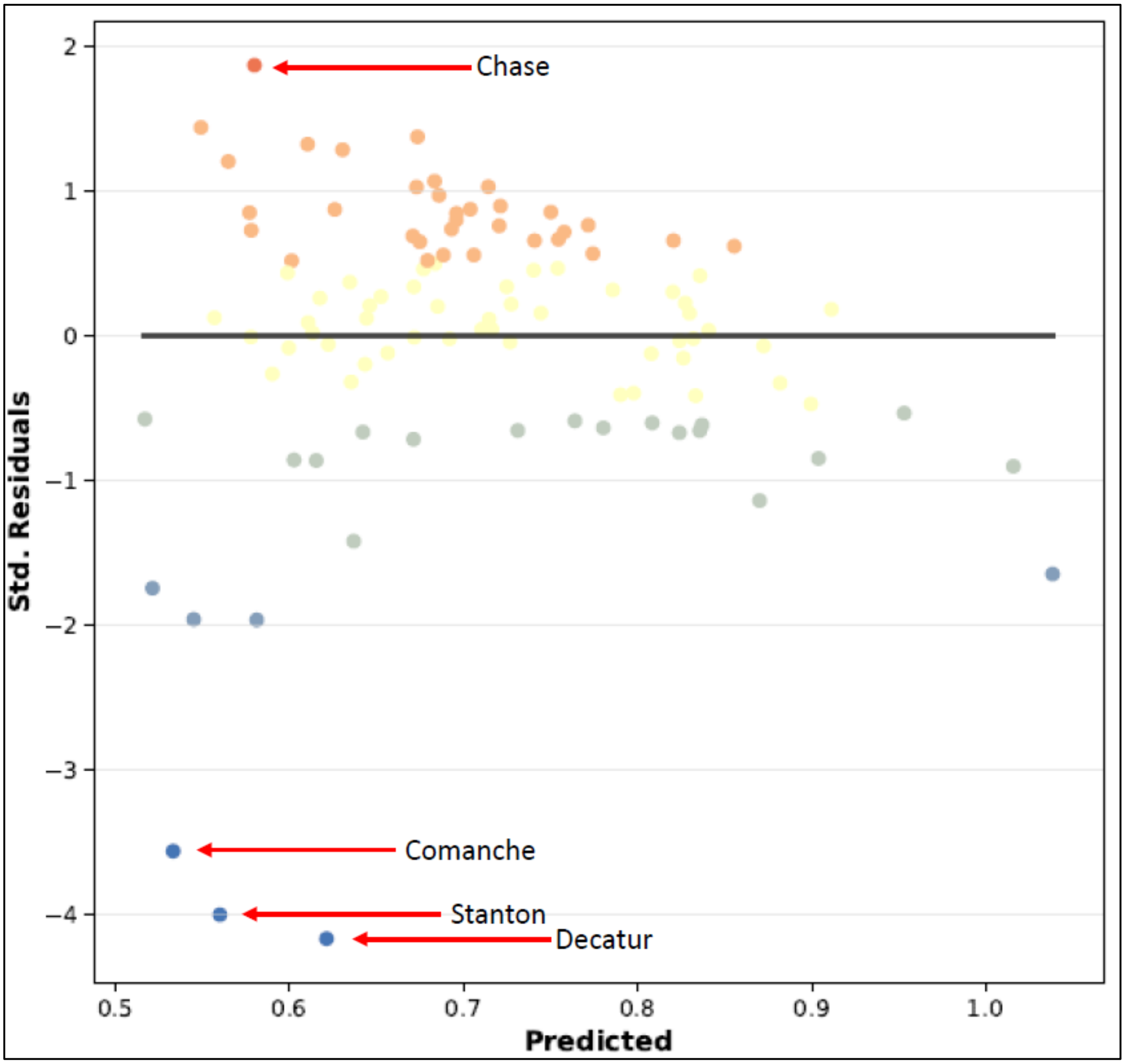


Figure 50. The OLS Residual vs. the Predicted Dependent Variable

OLS regression models the relationships between dependent and independent variables precisely when they were consistent across the study area, but when these relationships were heterogeneous and nonstationary across the study area, the regression equation creates an average of the mixed relationships present. The dominant method that deals with the regional variation and eliminate their impacts is the Geographically Weighted Regression (GWR) regression model.

Geographically Weighted Regression (GWR)

The GWR is a local model that creates an equation for every county in the state. In other words, the OLS used every single county in Kansas to calibrate the resultant equation, but the GWR models the nonstationary relationships over the study area so that each county gets a separate OLS equation calibrated based on the neighboring counties while using the same explanatory variables applied in the OLS model. Therefore, the coefficients of the exploratory variables will be different for each county in the study area.

The GWR was applied similar to the OLS. The number of traffic crashes involving teen drivers (LNLOCRASH) variable was entered as the dependent variable and both miles of rural non-state in a county (NONSTATE_RD), and the average number of passenger cars (LNLOAVG_PC) as explanatory variables. In the model type field, continuous (Gaussian), was selected

The GWR tool produces an attribute table that contains coefficients, local R-Squared, residuals, and some other parameter. Each of these parameters could be mapped to visualize their impact on the study area. The coefficient of the average number of passenger cars is shown in Figure 51. The dark areas show where the coefficient values were large and these were the locations with have the strongest relationship between the number of passenger cars variable and the number of crashes involving teen drivers. This does not mean that there were more teen-related crashes or more passenger cars in these dark areas, but it means that changing the number of passenger cars there will have more impact on the number of crashes involving teen drivers.

The resultant map of the other coefficient (miles of rural non-state roads) is shown in Figure 52.

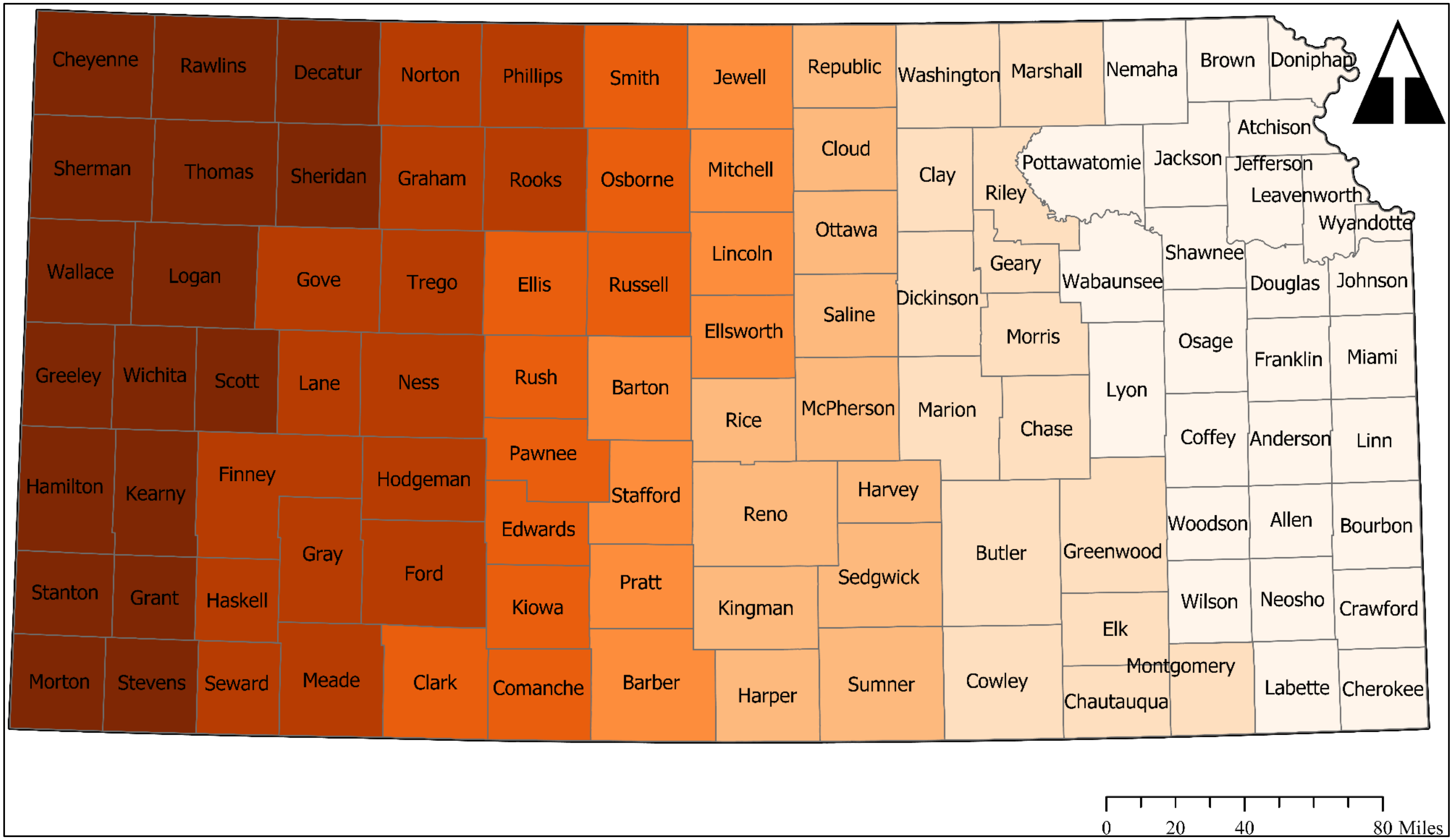


Figure 51. The Coefficient of Passenger Cars

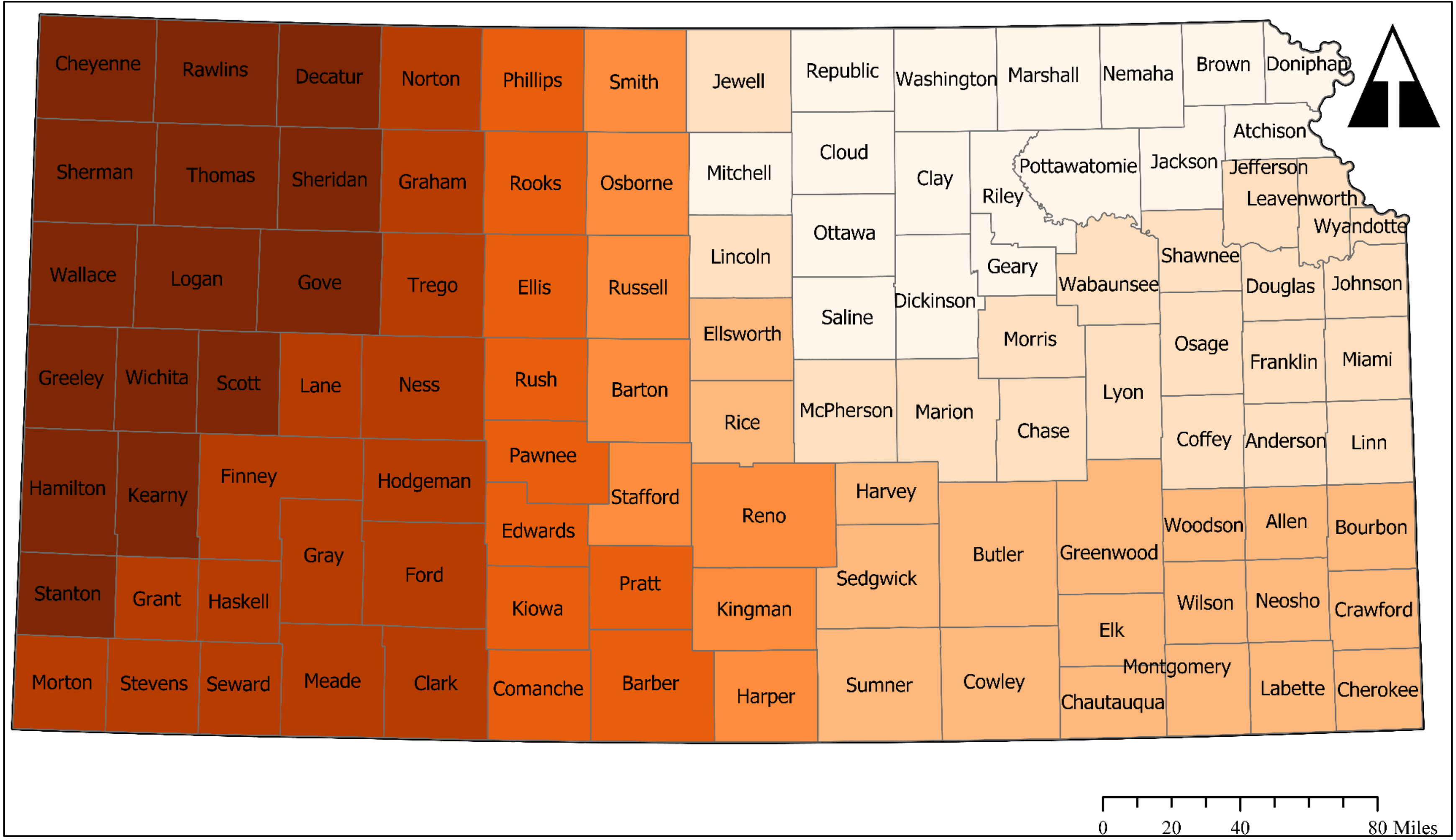


Figure 52. The Coefficient of Rural Non-State Roads

The dark areas show where the coefficient values were large and they represent a strong indicator for the number of crashes. Changing this exploratory variable in these locations will have more effect on the number of crashes involving teen drivers.

Prediction by Models

The OLS and GWR models were used to predict number of crashes involving teen drivers in 2026. Based on the available number of passenger cars and miles of non-state roads data from 2007 to 2016, the growth rate[1] of these variables for 2026 were calculated for each county. The resultant growth rates were used to predict the number of passenger cars and miles of non-state roads in 2026. Consequently, the predicted values were entered to the OLS and GWR models in order to predict the number of crashes involving teen drivers in 2026 for each county.

Based of the best available data, assuming nothing else changes in the state generally and counties specifically, the number of expected crashes based on the OLS model was predicted to be 10,824 statewide. However, the GWR model predicted the crash number in 2026 to be 10,795 crashes in the state. Given that the number of crashes involving teen drivers in 2016 was 11,172 (see Table 7), the OLS and GWR models predict a 3.11 percent and 3.37 percent crash reduction, respectively, if nothing else changed in the state. This predicted reduction is due to slight downward trends of growth rates in the two exploratory variables in the most populated counties (which also have the highest number of crashes) such as Douglas, Johnson, Leavenworth, Shawnee, and Wyandotte, as shown in Table 45. In the county level, the models predict the future trend of each county. This provides a useful indicator for related parties to identify where

[1] The statewide average growth rates for the number of passenger cars and miles of non-state roads were 0.002 percent and 0.047 percent, respectively.

they need to target their resources. For instance, the number of crashes that occurred in Shawnee County in 2016 was 860 crashes and the GWR model predicted that this number will increase to 983 (14.3%) by 2026. The details of predicted number of crashes for each county for the OLS and GWR models are shown in APPENDIX E (Table 43 and Table 44).

Validation of Models

Validation is a significant process to test the performance of model prediction when applied to an independent dataset that was not used in the modeling. The independent dataset used was the number of crashes involving teen driver, the number of registered passenger cars, and miles of non-state roads in each county of Kansas in 2017. The number of registered passenger cars, and miles of non-state roads dataset was entered into the models as exploratory variable, to predict the number of crashes involving teen drivers in 2017 in each county. Accordingly, the predicted numbers were compared to the real number of crashes. The validation step was performed for both OLS and GWR models.

For the OLS model the intercept, coefficient of miles of non-state roads, and coefficient of number of registered passenger cars were fixed for all counties, as they were: -1.065522, 0.019208, and 1.805782, respectively, but residuals was different based on counties. The resultant prediction number of crashes is shown in APPENDIX E, Table 46. As expected, the results were overestimated for some counties and underestimated for others. The overall predicted number of crashes was underestimated by 3.66 percent (411 crashes). That is_ the total number of predicted crashes was 10,801 crashes, whereas number of crashes involving teen drivers that occurred in Kansas in 2017 was 11,212 crashes.

However, the GWR's prediction number of crashes was overall better than the OLS's prediction number. The total predicted number by the GWR model was 10,883 crashes, which means it underestimated the crashes by 2.94 percent (329 crashes). Since each county in the GWR model had its own equation, intercept, coefficients, and residuals they were used separately to predict the number of crashes in each county. The details and results were displayed in Table 47.

At the county level, the model estimation for the number of crashes involving teen drivers is shown in Table 27. The table shows that the prediction of the OLS and GWR models were off by less than one percent for six counties, underestimated for four counties and overestimated for two counties. However, for 13 or14 counties (depending on the model used), the estimated number of crashes was off by more than 50 percent. The reason of these differences between predicted and actual number of crashes is not clear, and it could be caused by different factors, such as unusual weather or traffic patterns in those counties in 2017 compared to 2010-2016. An unusually high level of roadway construction or some other one-time event could also have been a factor.

Table 27. The Number of Counties Underestimated or Overstimated for Models

Percentage	**OLS**		**GWR**	
	Underestimated	Overestimated	Underestimated	Overestimated
< 1%	4	2	4	2
(1-4.9)%	3	6	3	6
(5-9.9)%	16	9	16	9
(10- 24.9)%	18	21	17	21
(25- 49.9)%	6	6	8	6
> 50%	4	10	3	10
Total	51	54	51	54

However, the counties that had been overestimated or underestimated by more than 25 percent were generally counties that had a low number of crashes. When the predicted number was off by a few crashes, the percentage of variance increased dramatically. For instance, the occurred number of crashes in Rawlins County was five crashes in 2017 while the predicted number of crashes was 7.65 crashes, which means it was overestimated by 53 percent, but the numerical difference between the actual number and the predicted number was only 2.65.

Furthermore, among the 30 counties that had the highest number of crashes (see Figure 26), only two counties (Jefferson and Wyandotte) had the predicted number of crashes off by more than 25 percent. It was not clear why the predicted number of crashes in Jefferson County was off by about 26 percent. Further analysis on Wyandotte County revealed that only 5.89 miles of non-state roads were reported in the list or county roadway miles provided by KDOT, but a brief review of the county's map (KDOT, 2015b) revealed that there are many more miles, which clearly shows that there is an error in the dataset for the non-state miles. If the correct number were available, it is believed that the predicted number of crashes would be much closer to the actual number.

UNIFIED SCHOOL DISTRICT LEVEL

The same analysis that was conducted statewide could be broken down into smaller levels, as desired, such as KDOT districts, counties, school district levels, zip codes, etc. In order to explain the implementation of this methodology in smaller areas, one of the unified school district (Wichita USD 259) was selected. USD 259 is headquartered in Wichita and covers an estimated area of 152.3 sq. miles. The major reason behind selecting USD 259 was this USD had the highest number of crashes involving teen drivers in the state during the study period. The number of fatal and injury crashes (Not PDOs) involving teen drivers in USD 259 between 2010 and 2016 was 3,019 crashes. Figure 53 shows all those crashes as red points spread out across

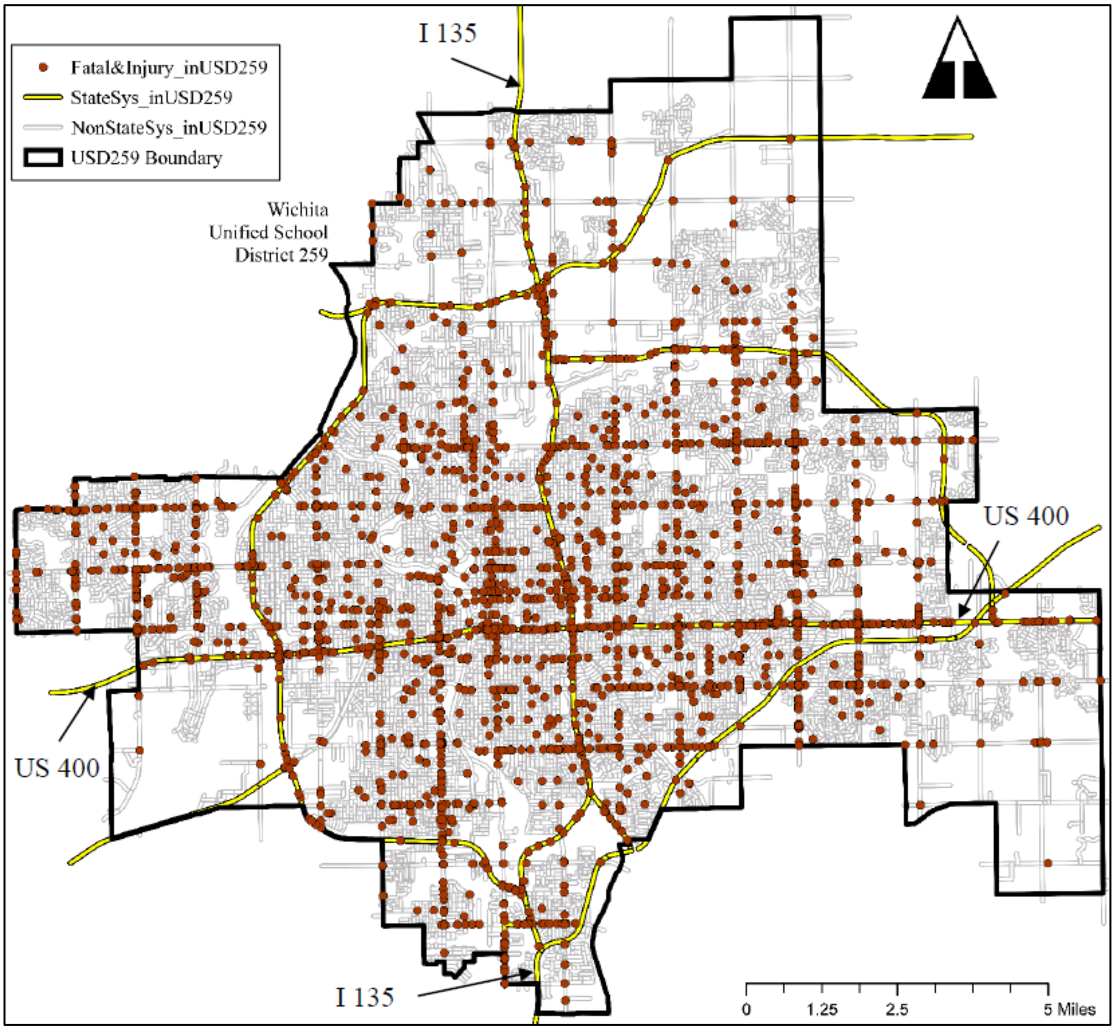


Figure 53. Fatal and Injury Crashes Involving Teen drivers in USD 259

the area. By looking at this dot map, it can be difficult to see any pattern. Therefore, more in-depth descriptive and statistical analysis was used to identify the existence of significant trends.

Teen drivers were involved in 22 fatal and 2,997 injury crashes in USD 259 during the study period, as shown in Table 28. The results showed different patterns of the number of crashes in terms of severity:

- The fatal crash type increased by 100 percent (doubled) from 2010 to 2016;
- The incapacitating crash type dropped 22.7 percent, which is the highest downward trend;
- The injury, not incapacitating type increased by 20 percent; and
- The possible injury type declined by 11.3 percent.

Table 28. The Severity of Teen Driver Involved in Crashes in USD 259 by Year

Injury Severity	2010	2011	2012	2013	2014	2015	2016	Total
Fatal Injury (K)	3	4	3	0	2	4	6	22
Incapacitating (A)	22	23	16	11	14	12	17	115
Injury, Not Incapacitating (B)	155	173	186	158	163	154	186	1,175
Possible Injury (C)	265	237	213	250	261	246	235	1,707
Total	445	437	418	419	440	416	444	3,019

The Central Mean and Standard Deviational Ellipse

The mean center of crashes for the study period was located almost in the center of the district. The standard deviational ellipse, shown as an orange ellipse in Figure 54, was nearly horizontal. The standard distances of the standard deviational ellipse were about 6.4 miles on the x-axis and 4.1 miles on the y-axis (see Figure 74 in APPENDIX C).

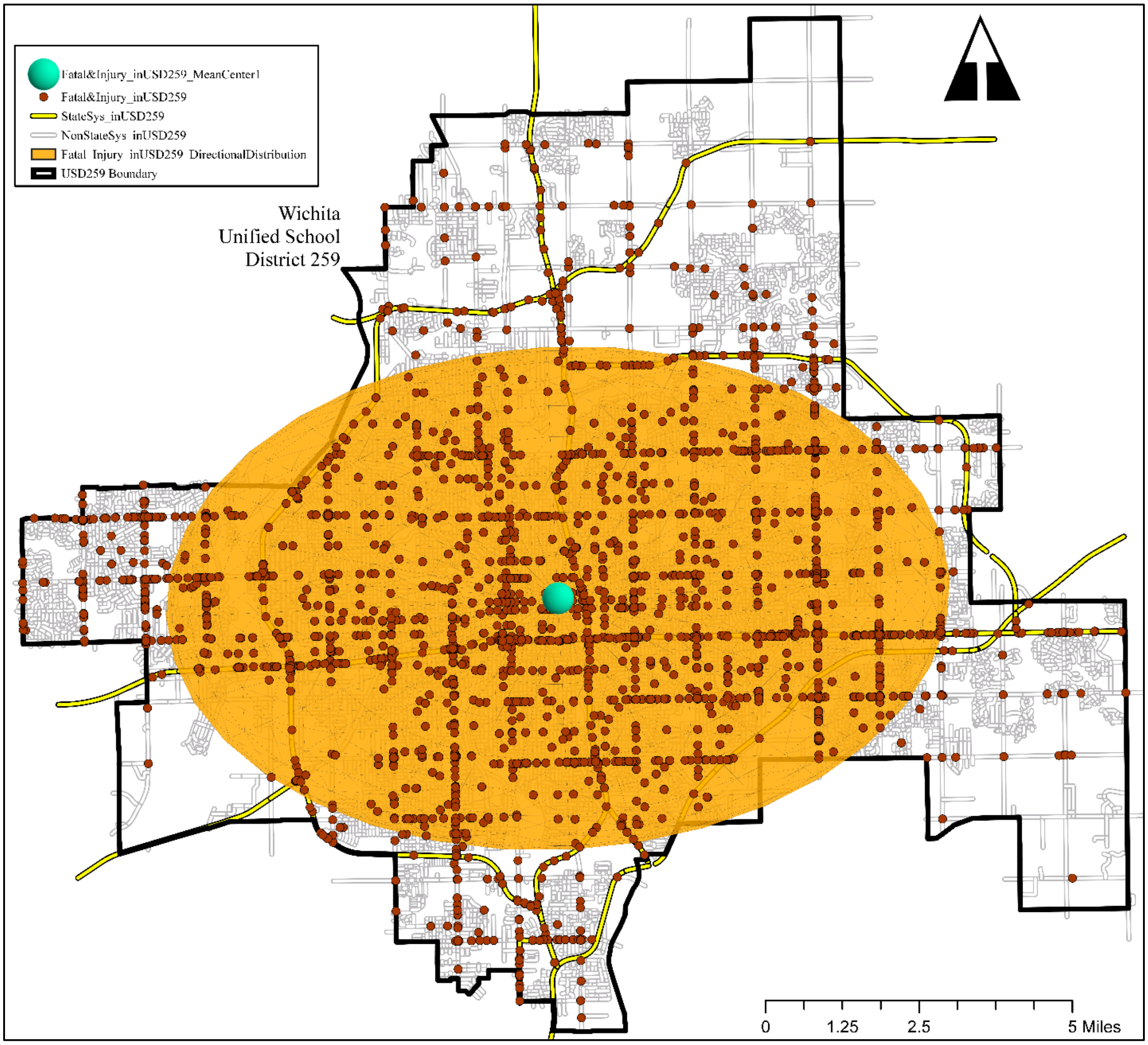
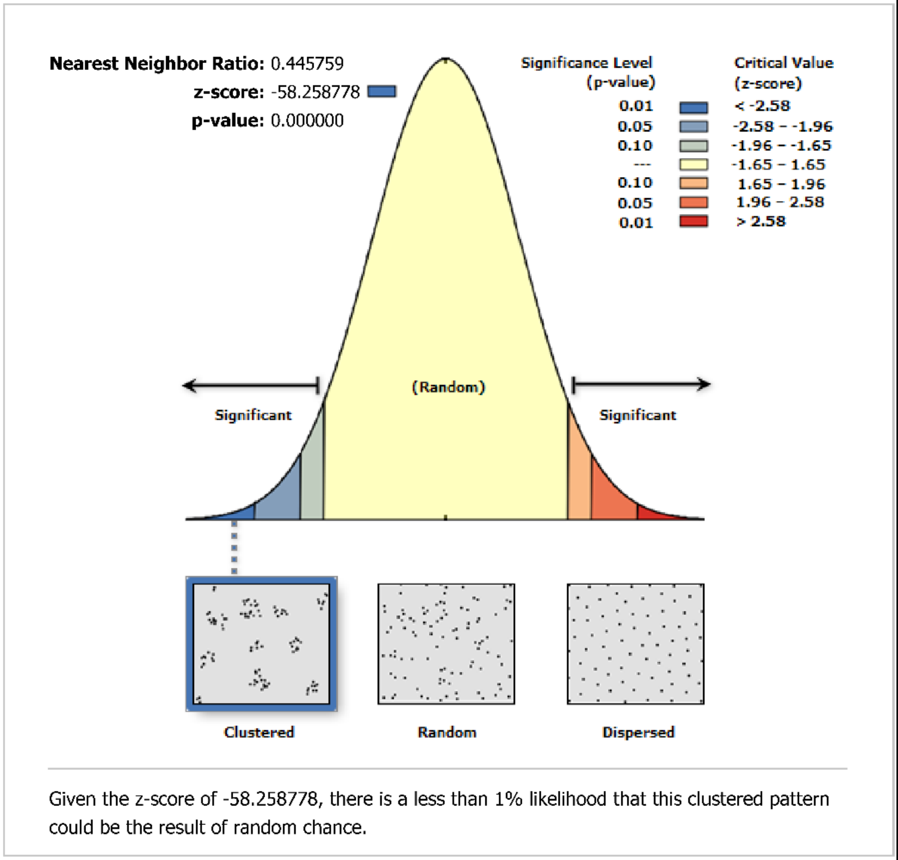


Figure 54. The Mean Center Standard Deviational Ellipse of USD 259

Average Nearest Neighbor

The Nearest Neighbor ratio for fatal and injury crashes involving teen drivers in the USD 259 was 0.446. Given the z-score of -58.259 and p-value of zero, there is no chance that this clustered pattern could be random, as shown in Figure 55.

Average Nearest Neighbor Summary



Average Nearest Neighbor Summary

Observed Mean Distance:	264.3076 US_Feet
Expected Mean Distance:	592.9387 US_Feet
Nearest Neighbor Ratio:	0.445759
z-score:	-58.258778
p-value:	0.000000

Figure 55. The Result of the Average Nearest Neighbor in the USD 259

Optimized Hot Spot Analysis

To extend the analysis, the study area was divided into identical polygons. Therefore, the "Count incidents within the hexagon grid" option was selected, which divides the district into several similar hexagonal polygons. In the "Bounding Polygons Defining Where Incidents Are Possible" field the shapefile of USD 259 boundary was uploaded to identify the border of the study area. The window of the Optimized Hotspot Analysis tool for USD 259 is shown in Figure 56.

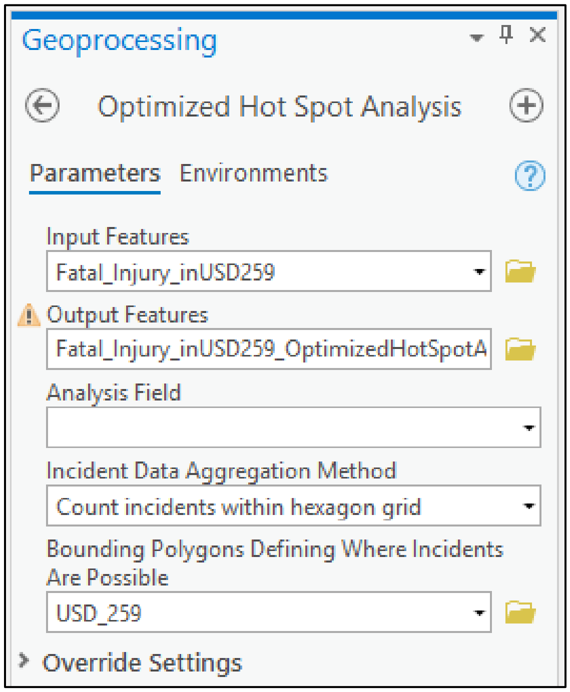


Figure 56. The Optimized Hot Spot Analysis Window for USD 259

The visualized results of the Optimized Hotspot Analysis for fatal and injury crashes involving teen drivers are shown in Figure 57. The map shows major hotspots with different levels of significance on route US 400 and I 135. This analysis validated all the 3,019 fatal and injury crashes, and it integrated them in 5,711 weighted hexagonal polygons. The size of the created hexagonal polygons was about 1,097 feet by 950 feet. The optimal fixed threshold distance based on peak clustering was found at 2,677 feet. In this threshold distance, none of the features had less than eight neighbors, which is a good indication for the accuracy of the results. Further details on the report are found in APPENDIX C, Figure 75.

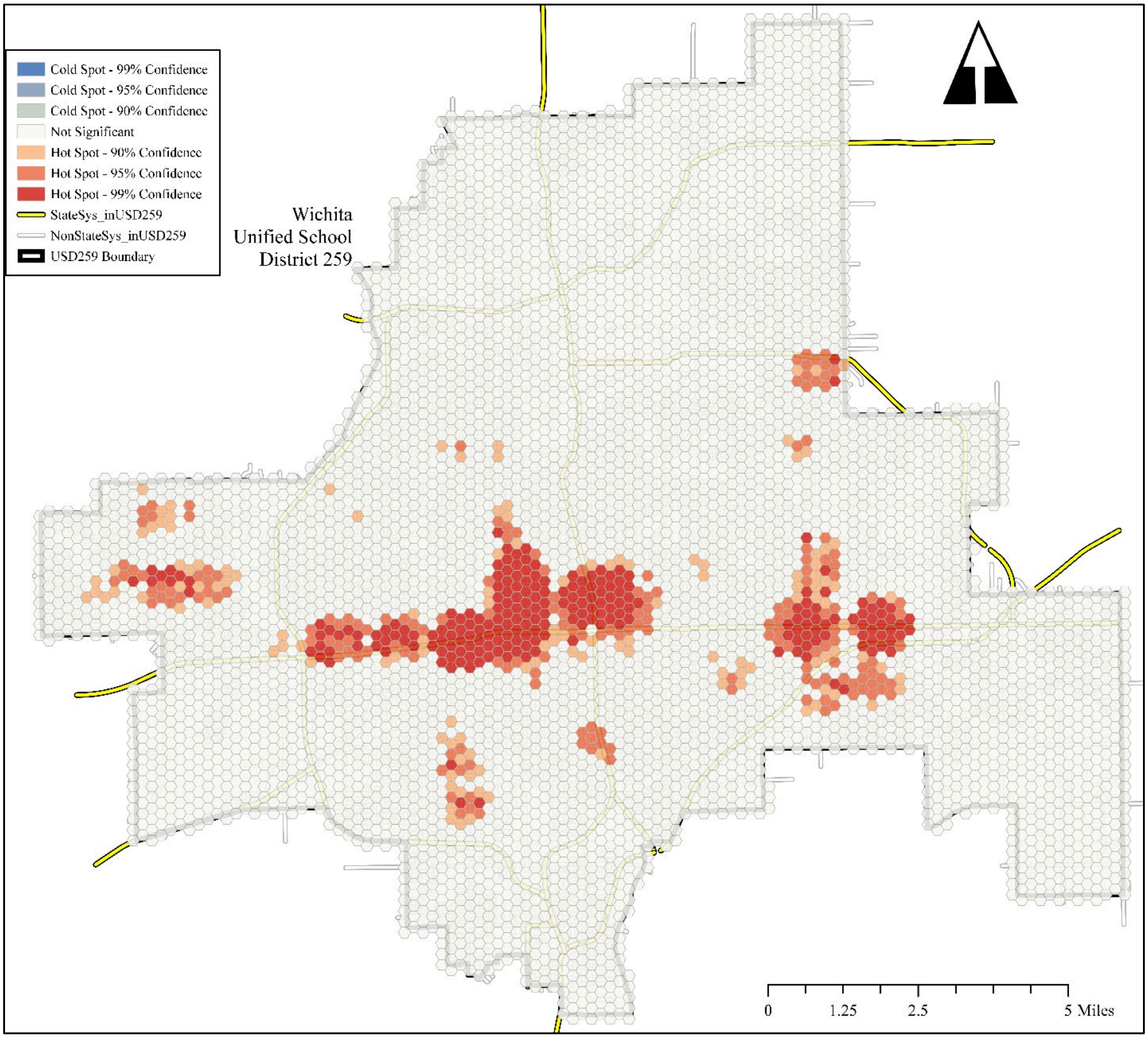


Figure 57. The Optimized Hot Spot Analysis Result in USD 259

Optimized Outlier Analysis

The cluster and outlier analysis results for fatal and injury crashes involving teen drivers and location of educational centers that contain teens in USD 259 are shown in Figure 58. From 1,002 statistically significant features, there were 186 outlier features that they were statistically significant. There were 40 red features that fall in the High-Low category of outliers spread out

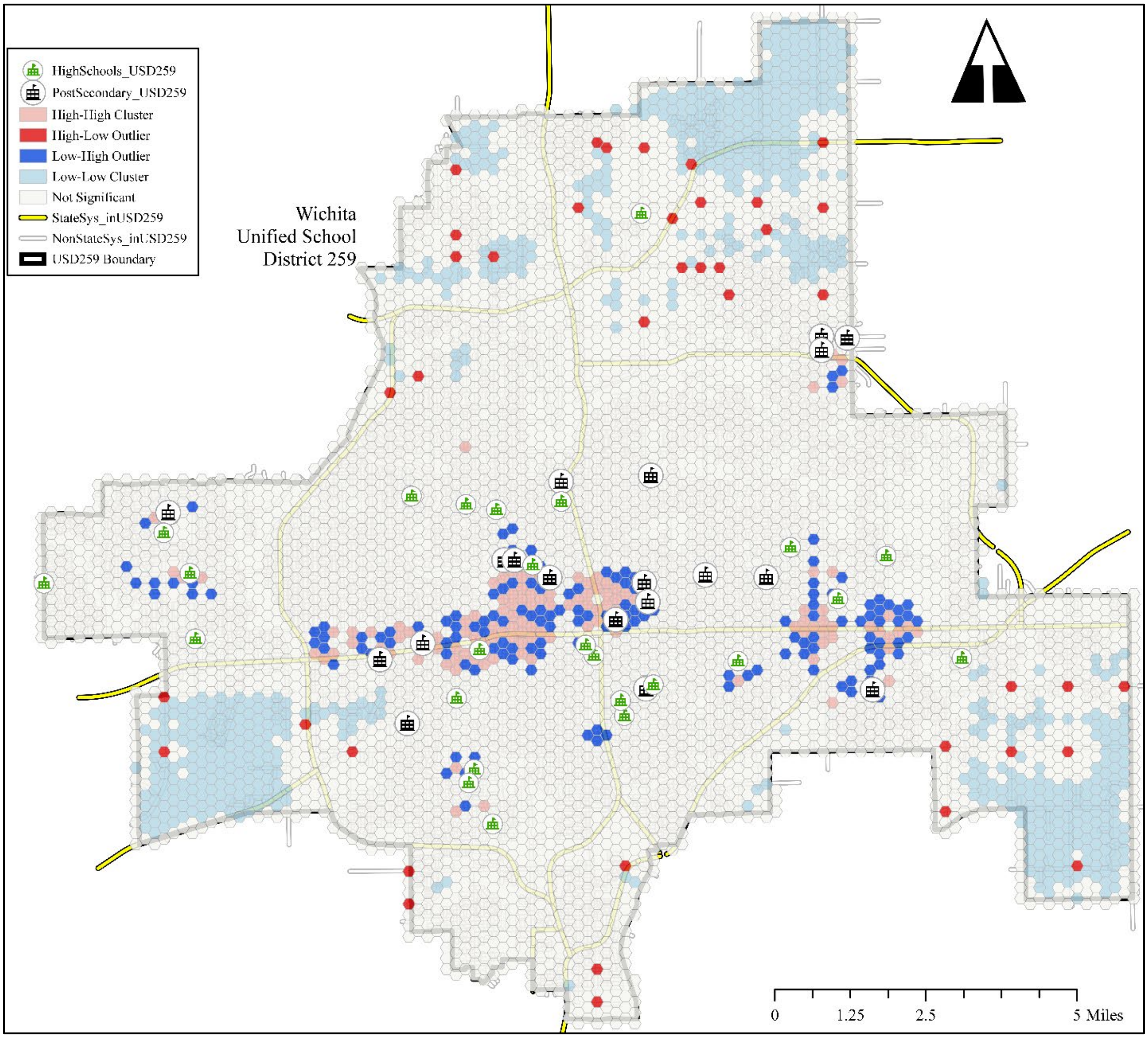


Figure 58. The Optimized Outlier Analysis Result for USD 259

in the north and south of the district. These were locations where fatal and injury crashes involving teen drivers were high, but the surrounding spots had a low or not statistically significant number of those crashes. The other outlier category is Low-High outliers, which are shown in dark blue polygons. There were 146 features that had a low number of fatal and injury crashes but their neighbors had a high or not statistically significant number of crashes.

In terms of clusters, 123 features were identified as High-High clustered features where the majority located on the route US 400 and I-135. On the other hand, 693 features were identified as Low-Low clustered features and most of them were located in the northern, south-eastern and south-western of the district. The High-High clusters mean that these features had a high number of fatal and injury crashes involving teen drivers and their surrounding features also had a statistically significant high number of fatal and injury crashes, and vice versa for the Low-Low clusters. The threshold distance based on peak clustering was found at 2,677 feet. In this threshold distance, none of the features had less than eight neighbors, which is a good indication for the accuracy of the results. Further details on the report could be found in APPENDIX C, Figure 76.

CONCLUDING REMARKS

The spatial analysis method can significantly help local agencies and KDOT to identify hotspots and high-low outliers to improve traffic safety in these locations. Teens usually frequent the educational centers. These centers, especially those located in or close to hotspots (see Figure 58), could be targeted by related parties to increase traffic safety. This is also a useful tool that promotes targeted enforcement campaigns and targeted public services messages like billboards. This research has been focused on teen driver-related crashes, but this analysis methodology could be easily applied to other crash categories. The next chapter will provide concluding remarks and discussion on ways this methodology can be expanded, and area of future work.

CHAPTER VI. FINDINGS AND RECOMMENDATIONS

This chapter includes a discussion of the meaning of the conducted analysis, as well as consideration of how the spatial analysis tools and resultant models benefit traffic safety. Furthermore, a discussion of the limitations that were faced during conducting this research is also given. Finally, further refinements and opportunities to expand this research in the future are listed.

- The utilized analyzing pattern toolset comprises significant functions that statistically measure clustered features and facilitated comparison of crash patterns for fatal and non-fatal teen-related crashes. By using the ANN tool, the existence of statistically significant clusters of teen-related crashes at the state level and the USD 259 level were identified. The Getis-Ord General G tool was used and it confirmed that the high values of fatal and non-fatal crashes were clustered. Moreover, Global Moran's I was used to measure spatial autocorrelation based on feature locations and attribute values and it was found that the data were clustered. At this stage, it was confirmed that fatal and non-fatal teen-related crashes were clustered and more investigation was needed to define their locations.
- The Optimized Hot Spot Analysis tool was used to identify statistically significant locations of high values clusters (hotspots) for crashes involving teen drivers across the state and in USD 259. Several locations were detected as hotspots, where teens had significantly more crashes than expected. The Optimized Cluster and Outlier Analysis was applied and several outliers were determined at the state level and in USD 259. The outliers indicate that, for these locations in the study area, the number of crashes were significantly different than the neighboring locations either positively or negatively.

- For identifying hotspot and outliers of teen driver-related crashes, typical hexagonal polygons were a better option to be used at the state level and local level than arbitrary geographical features such as districts, counties, zip codes, or blocks to avoid any bias by feature sizes.
- The OLS and GWR tools were used to determine the contributing factors behind observed spatial patterns of teen-related crashes and to predict the number of crashes involving teen drivers in each county in Kansas. Among 18 related exploratory variables that were prepared for modeling, only two were found to be statistically significant, and were used to build the predictive OLS and GWR models. The two exploratory variables were the number of miles of non-state roads in a county and the number of passenger cars in a county. With OLS a single model was built to represent the entire state, while with a separate GWR model were created for each county in the state.
- OLS and GWR models were used to predict the number of crashes involving teen drivers in the future for each county based on the growth rates of the exploratory variables. Assuming that no other global changes happened which could influence the number of the teen-related crashes, the models predicted a three percent reduction in the number of crashes, statewide by 2026.
- This research provides a useful indicator for related parties to identify where they can target their resources in order to improve teen driver safety. For instance, the number of crashes that occurred in Shawnee County in 2016 was 860 crashes and the GWR model predicted that this number will increase to 983 (an increase of 14.3 percent) by 2026 if nothing else was changed.

- The optimized hotspot analysis tool and the optimized clusters and outliers tool revealed the locations of hotspots and outliers in USD 259. This information can be used for crash reduction efforts such as targeted enforcement and education campaigns like billboards or other related messaging to increase traffic safety in the area.
- Even though statewide efforts to reduce teen-related crashes have been attempted (based on nonspatial factors such as gender, age, DUI, the presence of passengers, and distractions), it appears that overall teen crashes have begun to increase in the past few years in Kansas. The spatial analysis technique offers significant tools to better investigate and understand how teen-related crashes are statistically correlated and patterned and what associated factors are behind the patterns, providing additional information for decision-makers in mitigating these crashes.
- The methodology of the spatial analysis could be developed to an interactive interface by the state so the local agencies can get access and recognize statistically significant hot spots and outliers in their regions online.
- This research methodology was useful for analyzing a subset of crashes involving teen drivers; it can also be used to analyze other subsets such as alcohol-related crashes, older driver crashes, and commercial vehicle crashes or could be used for the entire statewide crash database.

LIMITATIONS

During the performance of this research, some limitations became evident, that if addressed could improve the utility of future research of this type.

- There were no data that showed how many miles teen drivers drive in a given year or the percentage of total vehicle-miles travelled by teen drivers. If such data were available, the models developed in this or future research could be improved.
- Useful data existed on the number of passenger cars at the county level. However, there were no data on how many of those passenger cars were driven by teens. These kinds of data could make a good predictor for the models if they were available. Therefore, the models were built on the best available data.
- There were good data on the population of teens in high schools and higher educational centers, but there was no information on how many of them had a driver license. The data on each of the population of teens and licensed teens came from separate databases, and it was beyond the scope of this dissertation to examine this relationship. However, if these databases were combined, it is believed that better models could be developed for examining teen-related crashes.
- The quality of the data used revealed some limitations. For instance, it was noticed that there was a small number of vehicles involved in teen-related crashes from 1900, 1909, 1937, etc., which were probably inputted incorrectly. Another example of this kind of error was the number of miles of non-state roads for Wyandotte County, which was only 5.89 miles. Additionally, the existing crash data were built based on the crashes that were reported to police, but some crashes such as PDOs are more likely to go unreported.

FUTURE STUDIES

Several avenues exist for future studies to extend the efforts of this research in order to expand the applications of the methodology and improve traffic safety.

- Applying this methodology for other age groups such as older drivers, or different crash categories such as commercial vehicles. The developed methodology can also analyze crashes on a specific type of roads (e.g., state highways or local roads), or investigate seasonal crash patterns such as crashes in sports events and holidays.
- Conducting temporal analysis using ArcGIS to compare crash rates between teen drivers and other age groups could be another interesting opportunity to understand how temporal factors impact their driving performance. For instance, in one of the secondary tasks in this research, the teen-related crash rates were compared between the dawn and dusk periods and day and night periods. The results showed that all types of crash rates were higher during dawn and dusk. Are these rates different for other age groups and/or for different types of crashes? Answering this question could be a worthy future study.
- The model underpredicted the number of crashes involving teen drivers in some counties and overrepresented the number of crashes in others. Chase County was a notable example of this. There is an opportunity to research the reported variance and to improve the methodology, which could result in more accurate prediction models. Additionally, an analysis could be performed to determine if a different model is needed for urban areas rather than using the statewide model.
- In this research, the Optimized Hot Spot Analysis tool and Optimized Cluster and Outlier Analysis tool in ArcGIS were used to identify hazardous locations for teen drivers statewide and for USD 259. Some researchers may desire to use the Network Screening

method presented in the HSM, which has the same ability. Additional research would be needed to determine if the network screening results would be similar to the results that obtained in this research, and to compare the findings of the current research.

REFERENCES

AASHTO, American Association of State Highway Transportation Official. 2010. "Highway Safety Manual." Washington, DC: AASHTO.

Aghajani, M.A., R.S. Dezfoulian, A.R. Arjroody, and M. Rezaei. 2017. 'Applying GIS to Identify the Spatial and Temporal Patterns of Road Accidents Using Spatial Statistics (Case Study: Ilam Province, Iran)', *Transportation Research Procedia*, 25: 2126-38.

Anderson, T. 2007. 'Comparison of Spatial Methods for Measuring Road Accident 'Hotspots': A Case Study of London', *Journal of Maps*, 3: 55-63.

Baker, S.P., L.-H. Chen, and G. Li. 2006. "National Evaluation of Graduated Driver Licensing Programs." 23. United States: United States. National Highway Traffic Safety Administration.

Bingham, C.R., J.S. Zakrajsek, F. Almani, J.T. Shope, and T.B. Sayer. 2015. 'Do as I Say, Not as I Do: Distracted Driving Behavior of Teens and Their Parents', *Journal of Safety Research*, 55: 21-29.

Block, A.W., and S. Walker. 2008. "2007 Motor Vehicle Occupant Safety Survey: Driver Education and Graduated Driver Licensing."

Borowsky, A., T. Oron-Gilad, and Y. Parmet. 2009. 'Age and Skill Differences in Classifying Hazardous Traffic Scenes', *Transportation Research Part F: Traffic Psychology and Behaviour*, 12: 277-87.

Borowsky, A., D. Shinar, and T. Oron-Gilad. 2010. 'Age, Skill, and Hazard Perception in Driving', *Accident Analysis & Prevention*, 42: 1240-49.

Bors, A.G., and N. Nasios. 2008. "Kernel Bandwidth Estimation in Methods Based on Probability Density Function Modelling." 1-4.

Braitman, K.A., B.B. Kirley, A.T. McCartt, and N.K. Chaudhary. 2008. 'Crashes of Novice Teenage Drivers: Characteristics and Contributing Factors', *Journal of Safety Research*, 39: 47-54.

Brunsdon, C., A.S. Fotheringham, and M.E. Charlton. 1996. 'Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity', *Geographical Analysis*, 28: 281-98.

Cammisa, M.X., A.F. Williams, and W.A. Leaf. 1999. 'Vehicles Driven by Teenagers in Four States', *Journal of Safety Research*, 30: 25-30.

Carney, C., K.K. Harland, and D.V. McGehee. 2016. 'Using Event-Triggered Naturalistic Data to Examine the Prevalence of Teen Driver Distractions in Rear-End Crashes', *Journal of Safety Research*, 57: 47-52.

Carter, P.M., C.R. Bingham, J.S. Zakrajsek, J.T. Shope, and T.B. Sayer. 2014. 'Social Norms and Risk Perception: Predictors of Distracted Driving Behavior among Novice Adolescent Drivers', *Journal of Adolescent Health*, 54: S32-S41.

CDC. 2017. 'Teen Drivers: Get the Facts 2015', Accessed October 16, 2017. https://www.cdc.gov/motorvehiclesafety/teen_drivers/teendrivers_factsheet.html.

Cestac, J., F. Paran, and P. Delhomme. 2011. 'Young Drivers' Sensation Seeking, Subjective Norms, and Perceived Behavioral Control and Their Roles in Predicting Speeding Intention: How Risk-Taking Motivations Evolve with Gender and Driving Experience', *Safety Science*, 49: 424-32.

Chen, L.-H., S.P. Baker, E.R. Braver, and G. Li. 2000. 'Carrying Passengers as a Risk Factor for Crashes Fatal to 16-and 17-Year-Old Drivers', *JAMA*, 283: 1578-82.

Compton, C.P. 2005. 'Injury Severity Codes: A Comparison of Police Injury Codes and Medical Outcomes as Determined by NASS CDS Investigators', *Journal of Safety Research*, 36: 483-84.

Crundall, D., P. Chapman, S. Trawley, L. Collins, E. Van Loon, B. Andrews, and G. Underwood. 2012. 'Some Hazards Are More Attractive Than Others: Drivers of Varying Experience Respond Differently to Different Types of Hazard', *Accident Analysis & Prevention*, 45: 600-09.

Curry, A.E., J. Hafetz, M.J. Kallan, F.K. Winston, and D.R. Durbin. 2011. 'Prevalence of Teen Driver Errors Leading to Serious Motor Vehicle Crashes', *Accident Analysis & Prevention*, 43: 1285-90.

Curry, A.E., K.B. Metzger, A.F. Williams, and B.C. Tefft. 2017. 'Comparison of Older and Younger Novice Driver Crash Rates: Informing the Need for Extended Graduated Driver Licensing Restrictions', *Accident Analysis & Prevention*, 108: 66-73.

Curry, A.E., J.H. Mirman, M.J. Kallan, F.K. Winston, and D.R. Durbin. 2012. 'Peer Passengers: How Do They Affect Teen Crashes?', *Journal of Adolescent Health*, 50: 588-94.

Danese, M., M. Lazzari, and B. Murgante. 2008. "Kernel Density Estimation Methods for a Geostatistical Approach in Seismic Risk Analysis: The Case Study of Potenza Hilltop Town (Southern Italy)." 415-29. Springer.

Danner, F., and B. Phillips. 2008. 'Adolescent Sleep, School Start Times, and Teen Motor Vehicle Crashes', *Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine*, 4: 533.

Delhomme, P., N. Chaurand, and F. Paran. 2012. 'Personality Predictors of Speeding in Young Drivers: Anger Vs. Sensation Seeking', *Transportation Research Part F: Traffic Psychology and Behaviour*, 15: 654-66.

DePesa, C., T. Raybould, S. Hurwitz, J. Lee, A. Gervasini, G.C. Velmahos, P.T. Masiakos, and H.M. Kaafarani. 2017. 'The Impact of the 2007 Graduated Driver Licensing Law in Massachusetts on the Rate of Citations and Licensing in Teenage Drivers', *Journal of Safety Research*, 61: 199-204.

Dingus, T.A., F. Guo, S. Lee, J.F. Antin, M. Perez, M. Buchanan-King, and J. Hankey. 2016. 'Driver Crash Risk Factors and Prevalence Evaluation Using Naturalistic Driving Data', *Proceedings of the National Academy of Sciences*, 113: 2636-41.

DMV. 2018a. 'Applying for a New License (Teen Drivers) in Kansas', Accessed December 6, 2018. https://www.dmv.org/ks-kansas/teen-drivers.php.

DMV. 2018b. 'Three Types of Driving Distractions', DMV, Accessed June 03, 2018. https://www.dmv.org/distracted-driving/three-types-of-distractions.php.

Duong, T. 2017. 'Ks: Kernel Density Estimation for Bivariate Data'. Accessed March 25, 2017. https://cran.r-project.org/web/packages/ks/vignettes/kde.pdf

Ehsani, J.P., C.R. Bingham, and J.T. Shope. 2013a. 'The Effect of the Learner License Graduated Driver Licensing Components on Teen Drivers' Crashes', *Accident Analysis & Prevention*, 59: 327-36.

Ehsani, J.P., C.R. Bingham, and J.T. Shope. 2013b. 'Graduated Driver Licensing for New Drivers', *American Journal of Preventive Medicine*, 45: 9-18.

Eichelberger, A.H., E.R. Teoh, and A.T. McCartt. 2015. 'Vehicle Choices for Teenage Drivers: A National Survey of Us Parents', *Journal of Safety Research*, 55: 1-5.

Esri. 2013. 'Hot Spot Analysis (Getis-Ord Gi*)', Esri, Accessed March 21, 2018. http://resources.arcgis.com/en/help/main/10.1/index.html#//005p00000010000000.

Esri. 2014a. 'How Cluster and Outlier Analysis (Anselin Local Moran's I) Works', Esri, Accessed December 20, 2018.

https://resources.arcgis.com/en/help/main/10.2/index.html#/How_Cluster_and_Outlier_Analysis_Anselin_Local_Moran_s_I_works/005p00000012000000/.

Esri. 2014b. 'How Directional Distribution (Standard Deviational Ellipse) Works', Esri, Accessed May 25, 2018. https://resources.arcgis.com/en/help/main/10.2/index.html#/How_Directional_Distribution_Standard_Deviational_Ellipse_works/005p0000001q000000/.

Esri. 2014c. 'How Hot Spot Analysis (Getis-Ord Gi*) Works', Esri, Accessed December 20, 2018. https://resources.arcgis.com/en/help/main/10.2/index.html#//005p00000011000000.

Esri. 2014d. 'How Mean Center Works', Esri, Accessed May 25, 2018. https://resources.arcgis.com/en/help/main/10.2/index.html#/How_Mean_Center_works/005p0000001s000000/.

Esri. 2014e. 'How Multi-Distance Spatial Cluster Analysis (Ripley's K-Function) Works', Esri, Accessed May 25, 2018. https://resources.arcgis.com/en/help/main/10.2/index.html#/How_Multi_Distance_Spatial_Cluster_Analysis_Ripley_s_K_function_works/005p0000000s000000/.

Esri. 2014f. 'An Overview of the Analyzing Patterns Toolset', Esri, Accessed May 25, 2018. https://resources.arcgis.com/en/help/main/10.2/index.html#/An_overview_of_the_Analyzing_Patterns_toolset/005p00000007000000/.

Esri. 2018a. 'Getting Started with Spatial Analysis', Esri.com, Accessed January 17, 2019. https://www.esri.com/training/catalog/5bca58f8f77b99238f845e1c/getting-started-with-spatial-analysis/.

Esri. 2018b. 'How Kernel Density Works', pro.arcgis.com, Accessed July 26, 2018. http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/how-kernel-density-works.htm.

Esri Events. 2017a. 'Spatial Data Mining I: Essentials of Cluster Analysis', Esri Events, Accessed May 20, 2017. https://www.youtube.com/watch?v=qQNOlfOYtyw&index=51&list=PLaPDDLTCmy4YcXpv_ypX3YicMHVUOuGYR.

Esri Events. 2017b. 'Using Spatial Statistics to Do More: Simple Approaches', Esri Events, Accessed May 20, 2017. https://www.esri.com/videos/watch?videoid=3d_8nQpSCgE.

Esri Events. 2018. 'Beyond Where: Modeling Spatial Relationships and Making Predictions', Esri Events, Accessed April 20, 2019. https://www.youtube.com/watch?v=5_tbFFeYXWM.

Evans, L. 2004. *Traffic Safety* (Science Serving Society: USA).

FHWA, Federal Highway Administration. 2018. 'National Highway System', Federal Highway Administration, Accessed January 20, 2019. https://www.fhwa.dot.gov/planning/national_highway_system/interstate_highway_system/routefinder/table01.cfm.

Fischer, M.M., and A. Getis. 2009. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* (Springer Science & Business Media).

Foss, R.D., and A.H. Goodwin. 2014. 'Distracted Driver Behaviors and Distracting Conditions among Adolescent Drivers: Findings from a Naturalistic Driving Study', *Journal of Adolescent Health*, 54: S50-S60.

Foss, R.D., R.L. Smith, F. Shi, and N.P. O'Brien. 2015. "School Start Times and Teenage Driver Motor Vehicle Crashes." University of North Carolina, Chapel Hill.

Foss, R.D., and A.F. Williams. 2015. 'Adolescent Drivers: Fine-Tuning Our Understanding', *Journal of Adolescent Health*, 57: S1-S5.

Fotheringham, A., C. Brunsdon, and M. Charlton. 2002. 'Geographically Weighted Regression: The Analysis of Spatially Varying Relationships'.

Getis, A., P.A. Longley, M.F. Goodchild, D.J. Maguire, and D. Rhind. 2005. *Spatial Statistics* (John Wiley and Sons Inc: New York, United States of America:).

Getis, A., and J.K. Ord. 1992. 'The Analysis of Spatial Association by Use of Distance Statistics', *Geographical Analysis*, 24: 189-206.

Gibin, M., P. Longley, and P. Atkinson. 2007. "Kernel Density Estimation and Percent Volume Contours in General Practice Catchment Area Analysis in Urban Areas." 11-13. Citeseer.

Glassbrenner, D. 2012. "An Analysis of Recent Improvements to Vehicle Safety."

Goldzweig, I.A., R.S. Levine, D. Schlundt, R. Bradley, G.D. Jones, R.J. Zoorob, and O.J. Ekundayo. 2013. 'Improving Seat Belt Use among Teen Drivers: Findings from a Service-Learning Approach', *Accident Analysis & Prevention*, 59: 71-75.

Grembek, O., K. Kim, O.H. Kwon, J. Lee, H. Liu, M.J. Park, S. Washington, D. Ragland, and S.M. Madanat. 2012. 'Experimental Evaluation of the Continuous Risk Profile (Crp) Approach to the Current Caltrans Methodology for High Collision Concentration Location Identification'.

Gross, F., T. Harmon, M. Albee, S. Himes, R. Srinivasan, D. Carter, and M. Dugas. 2016. "Evaluation of Four Network Screening Performance Measures."

Guidoum, A.C. 2013. 'Kernel Estimator and Bandwidth Selection for Density and Its Derivatives', *The kedd Package, version*, 1.

Hakkert, A.S., and L. Braimaister. 2002. "The Uses of Exposure and Risk in Road Safety Studies."

Hashimoto, S., S. Yoshiki, R. Saeki, Y. Mimura, R. Ando, and S. Nanba. 2016. 'Development and Application of Traffic Accident Density Estimation Models Using Kernel Density Estimation', *Journal of Traffic and Transportation Engineering (English edition)*, 3: 262-70.

Heck, K.E., and K.C. Nathaniel. 2011. "Driving among Urban, Suburban and Rural Youth in California." Subramaniam, A., Heck, K., Carlos, R., et al., 11. University of California, Agriculture & Natural Resources.

Hellinga, L.A., A.T. McCartt, and E.R. Haire. 2007. 'Choice of Teenagers' Vehicles and Views on Vehicle Safety: Survey of Parents of Novice Teenage Drivers', *Journal of Safety Research*, 38: 707-13.

Hellinga, L.A., A.T. McCartt, and S. Mandavilli. 2007. 'Temporal Patterns of Crashes of 16-to 17-Year-Old Drivers in Fairfax County, Virginia', *Traffic Injury Prevention*, 8: 377-81.

Herbel, S., L. Laing, and C. McGovern. 2010. "Highway Safety Improvement Program (Hsip) Manual." 160. Washington, DC: FHWA.

Hersman, D.A., and M.R. Rosekind. 2017. 'The Road to Zero Deaths from Motor Vehicle Crashes', *JAMA internal medicine*, 177: 1717-18.

Hicks, A. 2009. *Discovering Spatial and Temporal Patterns of Traffic Accidents in Stillwater, Oklahoma* (Citeseer).

Hilton, J. 2006. "Race and Ethnicity in Fatal Motor Vehicle Traffic Crashes 1999-2004."

Huestegge, L., E.-M. Skottke, S. Anders, J. Müsseler, and G. Debus. 2010. 'The Development of Hazard Perception: Dissociation of Visual Orientation and Hazard Processing', *Transportation Research Part F: Traffic Psychology and Behaviour*, 13: 1-8.

IIHS, Insurance Institute for Highway Safety. 2016. 'Fatality Facts: Teenagers 2015', Accessed November 12, 2017. http://www.iihs.org/iihs/topics/t/teenagers/fatalityfacts/teenagers.

IIHS, Insurance Institute for Highway Safety. 2018. 'Teenagers', Accessed December 6, 2018. https://www.iihs.org/iihs/topics/laws/graduatedlicenseintro.

IPEDS, Integrated Postsecondary Education Data System. 2019. 'College Map', National Center for Education Statistics. https://nces.ed.gov/ipeds/CollegeMap/.

IPP, Injury Prevention Program. 2009. "Kansas Injury Prevention Program Data Book." Kansas Department of Health & Environment, Bureau of Health Promotion and Office of Injury Prevention & Disability Programs Injury in Kansas, 76. Bureau of Health Promotion.

IPSR, Institute for Policy & Social Research. 2017. "Population Density Classifications in Kansas, by County, 2017." Lawrence, Kansas: The University of Kansas.

Jermakian, J.S., S. Bao, M.L. Buonarosa, J.R. Sayer, and C.M. Farmer. 2017. 'Effects of an Integrated Collision Warning System on Teenage Driver Behavior', *Journal of Safety Research*, 61: 65-75.

Jewett, A., R.A. Shults, and G. Bhat. 2016. 'Parental Perceptions of Teen Driving: Restrictions, Worry and Influence', *Journal of Safety Research*, 59: 119-23.

Kaafarani, H.M., J. Lee, C. Cropano, Y. Chang, T. Raybould, E. Klein, A. Gervasini, L. Petrovick, C. DePesa, and C.A. Camargo Jr. 2015. 'The Impact and Sustainability of the Graduated Driver Licensing Program in Preventing Motor Vehicle Crashes in Massachusetts', *Journal of Trauma and Acute Care Surgery*, 78: 265-71.

Kansas Department of Revenue. 2016. "Kansas Driving Handbook." 102. Topeka, Kansas.

Kansas State Department of Education. 2019. 'Usd Districts by County for 2018-2019', Kansas State Department of Education, Accessed January 25, 2019. https://uapps.ksde.org/Directory_Rpts/default.aspx.

Kansas.gov. 2018. 'The Sunflower State', Kansas.gov, Accessed December 29, 2018. https://portal.kansas.gov/.

KDOT, Kansas Department of Transportation. 2008. "Kansas Long Range Transportation Plan." Topeka, KS: KDOT.

KDOT, Kansas Department of Transportation. 2015a. "Kansas State Highway Strategic Safety Plan 2015." 104.

KDOT, Kansas Department of Transportation. 2019. 'Kansas Long Range Transportation Plan', KDOT, Accessed January 25, 2019. https://www.ksdot.org/bureaus/offchiefcoun/openrecords.asp.

KDOT, Kansas Department of Transportation,. 2015b. 'General Highway Map of Wyandotte County', KDOT, Accessed May 14, 2019. http://wfs.ksdot.org/arcgis_web_adaptor/rest/directories/arcgisoutput/County/halfInch/WyandotteCounty.pdf.

Klauer, S.G., F. Guo, B.G. Simons-Morton, M.C. Ouimet, S.E. Lee, and T.A. Dingus. 2014. 'Distracted Driving and Risk of Road Crashes among Novice and Experienced Drivers', *New England Journal of Medicine*, 370: 54-59.

Krisp, J.M., and O. Špatenková. 2010. 'Kernel Density Estimations for Visual Analysis of Emergency Response Data.' *Geographic Information and Cartography for Risk and Crisis Management* (Springer).

Kumfer, W., H. Liu, D. Wu, D. Wei, and S. Sama. 2017. 'Development of a Supplementary Driver Education Tool for Teenage Drivers on Rural Roads', *Safety Science*, 98: 136-44.

Kwon, O.H., M.J. Park, H. Yeo, and K. Chung. 2013. 'Evaluating the Performance of Network Screening Methods for Detecting High Collision Concentration Locations on Highways', *Accident Analysis & Prevention*, 51: 141-49.

Lam, W.H., S. Wong, and H.K. Lo. 2009. *Transportation and Traffic Theory 2009: Golden Jubilee* (Springer).

Lee, J.D. 2007. 'Technology and Teen Drivers', *Journal of Safety Research*, 38: 203-13.

Li, X., X. Yan, and S. Wong. 2015. 'Effects of Fog, Driver Experience and Gender on Driving Behavior on S-Curved Road Segments', *Accident Analysis & Prevention*, 77: 91-104.

Liu, Y., S. Singh, and R. Subramanian. 2015. "Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States, 2010 and 2011."

Lonero, L., and D. Mayhew. 2010. 'Large-Scale Evaluation of Driver Education: Review of the Literature on Driver Education Evaluation 2010 Update', *Washington, DC: AAA Foundation for Traffic Safety*.

Longley, P.A. 2005. *Geographical Information Systems: Principles, Techniques, Management and Applications (Abridged Edition)* (John Wiley & Sons, Inc.).

Loo, B.P., and T.K. Anderson. 2015. *Spatial Analysis Methods of Road Traffic Collisions* (CRC Press).

Manepalli, U.R., G.H. Bham, and S. Kandada. 2011. "Evaluation of Hotspots Identification Using Kernel Density Estimation (K) and Getis-Ord (Gi*) on I-630."

Masten, S.V., R.D. Foss, and S.W. Marshall. 2011. 'Graduated Driver Licensing and Fatal Crashes Involving 16-to 19-Year-Old Drivers', *JAMA*, 306: 1098-103.

McCartt, A.T., and E.R. Teoh. 2015. 'Tracking Progress in Teenage Driver Crash Risk in the United States since the Advent of Graduated Driver Licensing Programs', *Journal of Safety Research*, 53: 1-9.

McCartt, A.T., E.R. Teoh, M. Fields, K.A. Braitman, and L.A. Hellinga. 2010. 'Graduated Licensing Laws and Fatal Crashes of Teenage Drivers: A National Study', *Traffic Injury Prevention*, 11: 240-48.

McDonald, C.C., A.E. Curry, V. Kandadai, M.S. Sommers, and F.K. Winston. 2014. 'Comparison of Teen and Adult Driver Crash Scenarios in a Nationally Representative Sample of Serious Crashes', *Accident Analysis & Prevention*, 72: 302-08.

McDonald, N.C., A.B. McGrane, E.A. Rodgman, R.L. Steiner, W.M. Palmer, and B.F. Lytle. 2015. 'Assessing Multimodal School Travel Safety in North Carolina', *Accident Analysis & Prevention*, 74: 126-32.

McGehee, D.V., M.L. Reyes, and C. Carney. 2013. 'Age Vs. Experience-Evaluation of a Video Feedback Intervention for Newly Licensed Teen Drivers'. Iowa. University of Iowa.

McKnight, A., and A. McKnight. 2003. 'Young Novice Drivers: Careless or Clueless?', *Accident; Analysis and Prevention*, 35: 921.

Mitchel, A. 1999. *The Esri Guide to GIS Analysis: Volume 1, Geographic Patterns & Relationships*.

Mitchel, A. 2005. 'The Esri Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics', *ESRI Guide to GIS analysis*.

Momeni, H. 2016. 'Identifying Effective Geometric and Traffic Factors to Predict Crashes at Horizontal Curve Sections', Thesis, Kansas State University.

Mueller, A., and L. Trick. 2012. 'Driving in Fog: The Effects of Driving Experience and Visibility on Speed Compensation and Hazard Avoidance', *Accident; Analysis and Prevention*, 48: 472-79.

Muilenburg, J.L., W.D. Johnson, S.L. Usdan, L. Annang, and D.L. Clayton. 2007. 'Prevalence of Impaired Driving Behaviors in a Diverse, Rural, Southern Middle School', *Accident Analysis & Prevention*, 39: 1080-87.

Nakaya, T., and K. Yano. 2010. 'Visualising Crime Clusters in a Space-Time Cube: An Exploratory Data-Analysis Approach Using Space-Time Kernel Density Estimation and Scan Statistics', *Transactions in GIS*, 14: 223-39.

Neyens, D.M., and L.N. Boyle. 2007. 'The Effect of Distractions on the Crash Types of Teenage Drivers', *Accident Analysis & Prevention*, 39: 206-12.

Neyens, D.M., and L.N. Boyle. 2008. 'The Influence of Driver Distraction on the Severity of Injuries Sustained by Teenage Drivers and Their Passengers', *Accident Analysis & Prevention*, 40: 254-59.

NHTSA. 2018a. 'Distracted Driving', NHTSA, Accessed June 3, 2018. https://www.nhtsa.gov/risky-driving/distracted-driving.

NHTSA, National Highway Traffic Safety Administration. 2013. "How Vehicle Age and Model Year Relate to Driver Injury Severity in Fatal Crashes." 825. NHTSA.

NHTSA, National Highway Traffic Safety Administration. 2015. "School Start Times and Teen Driver Crashes." Washington, DC: NHTSA.

NHTSA, National Highway Traffic Safety Administration. 2017. "2016 Fatal Motor Vehicle Crashes: Overview." 1-9. NHTSA.

NHTSA, National Highway Traffic Safety Administration. 2018b. "Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States, 2015." USDOT. NHTSA.

NHTSA, National Highway Traffic Safety Administration. 2018c. "Passenger Vehicle Occupant Injury Severity by Vehicle Age and Model Year in Fatal Crashes." Washington, DC: NHTSA.

NHTSA, National Highway Traffic Safety Administration. 2019. 'Fatality Analysis Reporting System (Fars)', United States Department of Transportation, Accessed January 26, 2019. https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars.

O'Neill, B., and D. Mohan. 2002. 'Reducing Motor Vehicle Crash Deaths and Injuries in Newly Motorising Countries', *BMJ: British Medical Journal*, 324: 1142.

OECD. 2006. "Young Drivers: The Road to Safety." 8. Paris, France: OECD Publishing.

Ord, J.K., and A. Getis. 1995. 'Local Spatial Autocorrelation Statistics: Distributional Issues and an Application', *Geographical Analysis*, 27: 286-306.

Oris, W.N. 2011. 'Spatial Analysis of Fatal Automobile Crashes in Kentucky'.

Ouimet, M.C., A.K. Pradhan, A. Brooks-Russell, J.P. Ehsani, D. Berbiche, and B.G. Simons-Morton. 2015. 'Young Drivers and Their Passengers: A Systematic Review of Epidemiological Studies on Crash Risk', *Journal of Adolescent Health*, 57: S24-S35. e6.

Ouimet, M.C., A.K. Pradhan, B.G. Simons-Morton, G. Divekar, H. Mehranian, and D.L. Fisher. 2013. 'The Effect of Male Teenage Passengers on Male Teenage Drivers: Findings from a Driving Simulator Study', *Accident Analysis & Prevention*, 58: 132-39.

Peden, M., R. Scurfield, D. Sleet, D. Mohan, A.A. Hyder, E. Jarawan, and C.D. Mathers. 2004. "World Report on Road Traffic Injury Prevention." World Health Organization Geneva.

Peek-Asa, C., C. Britton, T. Young, M. Pawlovich, and S. Falb. 2010. 'Teenage Driver Crash Incidence and Factors Influencing Crash Injury by Rurality', *Journal of Safety Research*, 41: 487-92.

Pimpler, E. 2017. *Spatial Analytics with ArcGIS* (Packt Publishing: UK).

Preusser, D.F., and J. Tison. 2007. 'GDL Then and Now', *Journal of Safety Research*, 38: 159-63.

Rhodes, N., and K. Pivik. 2011. 'Age and Gender Differences in Risky Driving: The Roles of Positive Affect and Risk Perception', *Accident Analysis & Prevention*, 43: 923-31.

Rodrigue, J.-P. 2017. 'The Geography of Transport Systems', Accessed November, 10. https://people.hofstra.edu/geotrans/eng/methods/ch1m4en.html.

Ryb, G.E., P.C. Dischinger, G. McGwin, and R.L. Griffin. 2011. "Crash-Related Mortality and Model Year: Are Newer Vehicles Safer?", 113. Association for the Advancement of Automotive Medicine.

Sagberg, F., and T. Bjørnskau. 2006. 'Hazard Perception and Driving Experience among Novice Drivers', *Accident Analysis & Prevention*, 38: 407-14.

Satria, R., and M. Castro. 2016. 'GIS Tools for Analyzing Accidents and Road Design: A Review', *Transportation Research Procedia*, 18: 242-47.

Scialfa, C.T., M.C. Deschênes, J. Ference, J. Boone, M.S. Horswill, and M. Wetton. 2011. 'A Hazard Perception Test for Novice Drivers', *Accident Analysis & Prevention*, 43: 204-08.

Scott-Parker, B., and O. Oviedo-Trespalacios. 2017. 'Young Driver Risky Behaviour and Predictors of Crash Risk in Australia, New Zealand and Colombia: Same but Different?', *Accident Analysis & Prevention*, 99: 30-38.

Scott, L., and M. Pratt. 2009. 'Answering Why Questions: An Introduction to Using Regression Analysis with Spatial Data', esri.com, Accessed January 15, 2018. https://www.esri.com/news/arcuser/0309/files/why.pdf.

Senserrick, T., and A. Williams. 2015. "Summary of Literature of the Effective Components of Graduated Driver Licensing Systems."

Shafabakhsh, G.A., A. Famili, and M.S. Bahadori. 2017. 'GIS-Based Spatial Analysis of Urban Traffic Accidents: Case Study in Mashhad, Iran', *Journal of Traffic and Transportation Engineering (English edition)*, 4: 290-99.

Shell, D.F., I.M. Newman, A.L. Córdova-Cazar, and J.M. Heese. 2015. 'Driver Education and Teen Crashes and Traffic Violations in the First Two Years of Driving in a Graduated Licensing System', *Accident Analysis & Prevention*, 82: 45-52.

Shinar, D. 2007. 'Traffic Safety and Human Behavior'.

Shinar, D. 2017. *Traffic Safety and Human Behavior* (Emerald Publishing Limited: New York).

Shope, J.T., and C.R. Bingham. 2008. 'Teen Driving', *American Journal of Preventive Medicine*, 35: S261-S71.

Shults, R.A. 2016. 'Graduated Driver Licensing Night Driving Restrictions and Drivers Aged 16 or 17 Years Involved in Fatal Night Crashes—United States, 2009–2014', *MMWR. Morbidity and mortality weekly report*, 65.

Shults, R.A., E. Olsen, and A.F. Williams. 2015. 'Driving among High School Students-United States, 2013', *MMWR. Morbidity and mortality weekly report*, 64: 313-17.

Shults, R.A., and A.F. Williams. 2017. 'Trends in Teen Driver Licensure, Driving Patterns and Crash Involvement in the United States, 2006–2015', *Journal of Safety Research*, 62: 181-84.

Simons-Morton, B.G., F. Guo, S.G. Klauer, J.P. Ehsani, and A.K. Pradhan. 2014. 'Keep Your Eyes on the Road: Young Driver Crash Risk Increases According to Duration of Distraction', *Journal of Adolescent Health*, 54: S61-S67.

Simpson, H. 1996. *New to the Road: Reducing the Risks for Young Motorists* (Youth Enhancement Service, Brain Information Service, UCLA School of Medicine).

Singh, S. 2015. "Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey."

Singh, S. 2018. "Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey." 3.

Sipos, T. 2017. 'Spatial Statistical Analysis of the Traffic Accidents', *Periodica Polytechnica Transportation Engineering*, 45: 101-05.

Songchitruksa, P., and X. Zeng. 2010. 'Getis-Ord Spatial Statistics to Identify Hot Spots by Using Incident Management Data', *Transportation Research Record: Journal of the Transportation Research Board*: 42-51.

Srinivasan, R., F. Gross, B. Lan, and G. Bahar. 2016. "Reliability of Safety Management Methods: Network Screening."

Stone, L.M., and C.W. Runyan. 2005. 'High School Off-Campus Lunch Policies and Adolescent Motor Vehicle Crash Risks', *Journal of Adolescent Health*, 36: 5-8.

Subramanian, R. 2005. 'Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States, 2001', *Young*, 1: 3.

Subramanian, R. 2012. "Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States, 2008 and 2009."

Taubman–Ben-Ari, O. 2014. 'How Are Meaning in Life and Family Aspects Associated with Teen Driving Behaviors?', *Transportation Research Part F: Traffic Psychology and Behaviour*, 24: 92-102.

Taubman–Ben-Ari, O., S. Kaplan, T. Lotan, and C.G. Prato. 2015. 'Parents' and Peers' Contribution to Risky Driving of Male Teen Drivers', *Accident Analysis & Prevention*, 78: 81-86.

Tefft, B.C., A.F. Williams, and J.G. Grabowski. 2014. 'Driver Licensing and Reasons for Delaying Licensure among Young Adults Ages 18-20, United States, 2012', *Injury Epidemiology*, 1: 4.

Time and Date. 2019. 'Sun Graph', timeanddate.com, Accessed March 21, 2019. https://www.timeanddate.com/sun/usa/topeka.

Tobler, W.R. 1970. 'A Computer Movie Simulating Urban Growth in the Detroit Region', *Economic Geography*, 46: 234-40.

Tractinsky, N., E.S. Ram, and D. Shinar. 2013. 'To Call or Not to Call—That Is the Question (While Driving)', *Accident Analysis & Prevention*, 56: 59-70.

U.S. Census Bureau. 2017. 'Tigerweb', U.S. Census Bureau, Accessed January 20, 2019. https://tigerweb.geo.census.gov/tigerwebmain/TIGERweb_apps.html.

U.S. Census Bureau. 2019a. '2013-2017 American Community Survey 5-Year Estimates', American FactFinder, Accessed January 19, 2019. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_B01003&prodType=table.

U.S. Census Bureau. 2019b. 'American Factfinder', U.S. Department of Commerce Accessed February 19, 2019. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml.

Vaca, F., H.G. Garrison, M.P. McKay, and C.S. Gotschall. 2006. 'National Highway Traffic Safety Administration (Nhtsa) Notes', *Annals of Emergency Medicine*, 48: 470.

Vorona, R.D., M. Szklo-Coxe, A. Wu, M. Dubik, Y. Zhao, and J.C. Ware. 2011. 'Dissimilar Teen Crash Rates in Two Neighboring Southeastern Virginia Cities with Different High School Start Times', *Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine*, 7: 145.

Wade, T., and S. Sommer. 2006. *A to Z GIS, an Illustrated Dictionary of Geographic Information Systems* (Esri Press).

Wallpaperama. 2006. 'List of All the 50 States That Make up the United States of America', wallpaperama.com, Accessed June 10, 2017. http://www.wallpaperama.com/forums/list-of-all-the-50-states-that-make-up-the-united-states-of-america-t390.html.

Webb, C. 2016. "Motor Vehicle Traffic Crashes as a Leading Cause of Death in the United States, 2012–2014."

Westlake, E.J., and L.N. Boyle. 2012. 'Perceptions of Driver Distraction among Teenage Drivers', *Transportation Research Part F: Traffic Psychology and Behaviour*, 15: 644-53.

Wetton, M.A., A. Hill, and M.S. Horswill. 2011. 'The Development and Validation of a Hazard Perception Test for Use in Driver Licensing', *Accident Analysis & Prevention*, 43: 1759-70.

WHO, World Health Organization. 2007. "Youth and Road Safety." Switzerland: World Health Organization.

WHO, World Health Organization. 2017. 'World Health Statistics 2017: Monitoring Health for the SDGs, Sustainable Development Goals'.

Williams, A.F. 2003. 'Teenage Drivers: Patterns of Risk', *Journal of Safety Research*, 34: 5-15.

Williams, A.F. 2017. 'Graduated Driver Licensing (GDL) in the United States in 2016: A Literature Review and Commentary', *Journal of Safety Research*, 63: 29-41.

Williams, A.F., W.A. Leaf, B.G. Simons-Morton, and J.L. Hartos. 2006. 'Vehicles Driven by Teenagers in Their First Year of Licensure', *Traffic Injury Prevention*, 7: 23-30.

Williams, A.F., A.T. McCartt, and L.B. Sims. 2016. 'History and Current Status of State Graduated Driver Licensing (GDL) Laws in the United States', *Journal of Safety Research*, 56: 9-15.

Williams, A.F., and R.A. Shults. 2010. 'Graduated Driver Licensing Research, 2007–Present: A Review and Commentary', *Journal of Safety Research*, 41: 77-84.

Williams, A.F., and B.C. Tefft. 2014. 'Characteristics of Teens-with-Teens Fatal Crashes in the United States, 2005–2010', *Journal of Safety Research*, 48: 37-42.

Wong, E., and R. Kunz. 2017. *Abbreviated Injury Scale* (Springer International Publishing: Cham).

Xie, Z., and J. Yan. 2008. 'Kernel Density Estimation of Traffic Accidents in a Network Space', *Computers, Environment and Urban Systems*, 32: 396-406.

Xu, C., W. Wang, and P. Liu. 2013. 'Identifying Crash-Prone Traffic Conditions under Different Weather on Freeways', *Journal of Safety Research*, 46: 135-44.

Young, J., and P.Y. Park. 2014. 'Hotzone Identification with GIS-Based Post-Network Screening Analysis', *Journal of Transport Geography*, 34: 106-20.

YOURS, Youth for Road Safety. 2012. "Youth and Road Safety Action Kit." Amsterdam: YOURS-Youth for Road Safety.

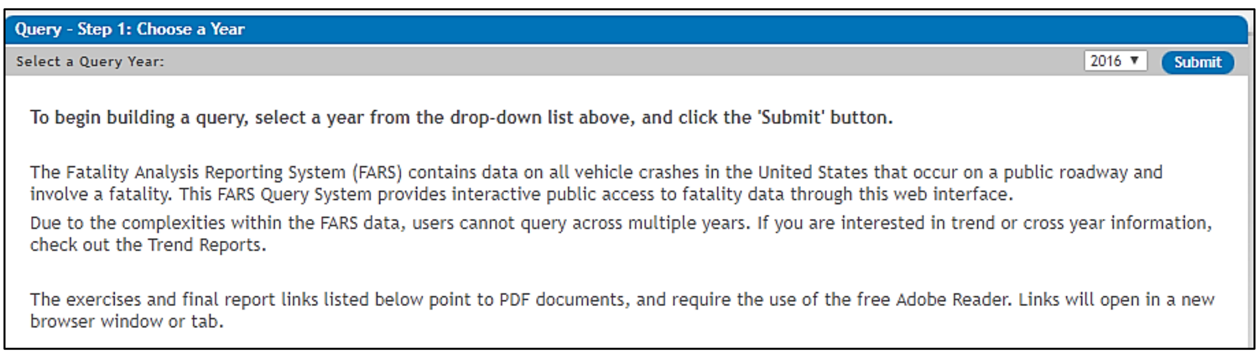
APPENDIX A

Example of Obtaining Data from FARS

In this example, the goal was extracting a spreadsheet of data on teen drivers involved in fatal crashes in 2016 in Kansas. From the FARS website below, the following simple steps were needed to download the data: (https://www-fars.nhtsa.dot.gov//QueryTool/QuerySection/SelectYear.aspx)

Step I

The data were downloaded for each year separately. In the first step, the query year (2016) was chosen from the drop-down list.



Step II

Within each years' dataset, there were three combinations of options categorized based on personal data, crash data, and vehicle data. Option 1 included information and variables about all people involved in each fatal crash such as age, gender, and location of occupants. Option 2 provided information about each fatal crash such as information on the location of crashes, lighting conditions, and weather conditions. The last option (Option 3) included information and variables related to the vehicles involved in each fatal crash such as type, year, and model of vehicles.

For the purpose of this example, the query option (Option 1) was selected.

Submit

Now select the combination of tables that contains the fields you are interested in. (In the next step, you will select the specific fields.)

Please click on the links to see the variables in the respective tables in a new window.

Option 1 (Crash / Person (Includes Occupants and Non Occupants))

Please select this option to query on variables that belongs to Crash, Person (Includes Occupant and Non Occupant) tables

Choose Option 1 if you want to include results for ALL PEOPLE (Occupants AND Non-Occupants) and /or CRASH Level information, but NOT VEHICLE or DRIVER Level information

Examples uses:

- You are looking for Alcohol or Drug test results for drivers and non-motorists involved in fatal crashes.
- You are looking for persons involved in fatal crashes that occurred at night.
- You are looking for the count of children involved in fatal crashes.

Option 2 (Crash / Non Occupant / Pedestrian / Bicyclist)

Please select this option to query on variables that belong to Crash, Non Occupant, Pedestrian and Bicyclist tables

Choose Option 2 if you want to include PERSON Level information only for Non-Occupants and/or CRASH Level information, but NOT VEHICLE or DRIVER Level information

Examples uses:

- You are looking for the location of bicyclists and their use of safety equipment when involved in a fatal crash.
- You are looking for common activities engaged in by pedestrians prior to being involved in a fatal crash.
- You are looking for a distribution of persons operating a personal conveyance or riding a bicycle by age group.

Option 3 (Crash / Vehicle / Driver / Precrash / Occupant)

Please select this option to query on variables that belongs to Crash, Vehicle, Driver, Precrash and Occupant tables

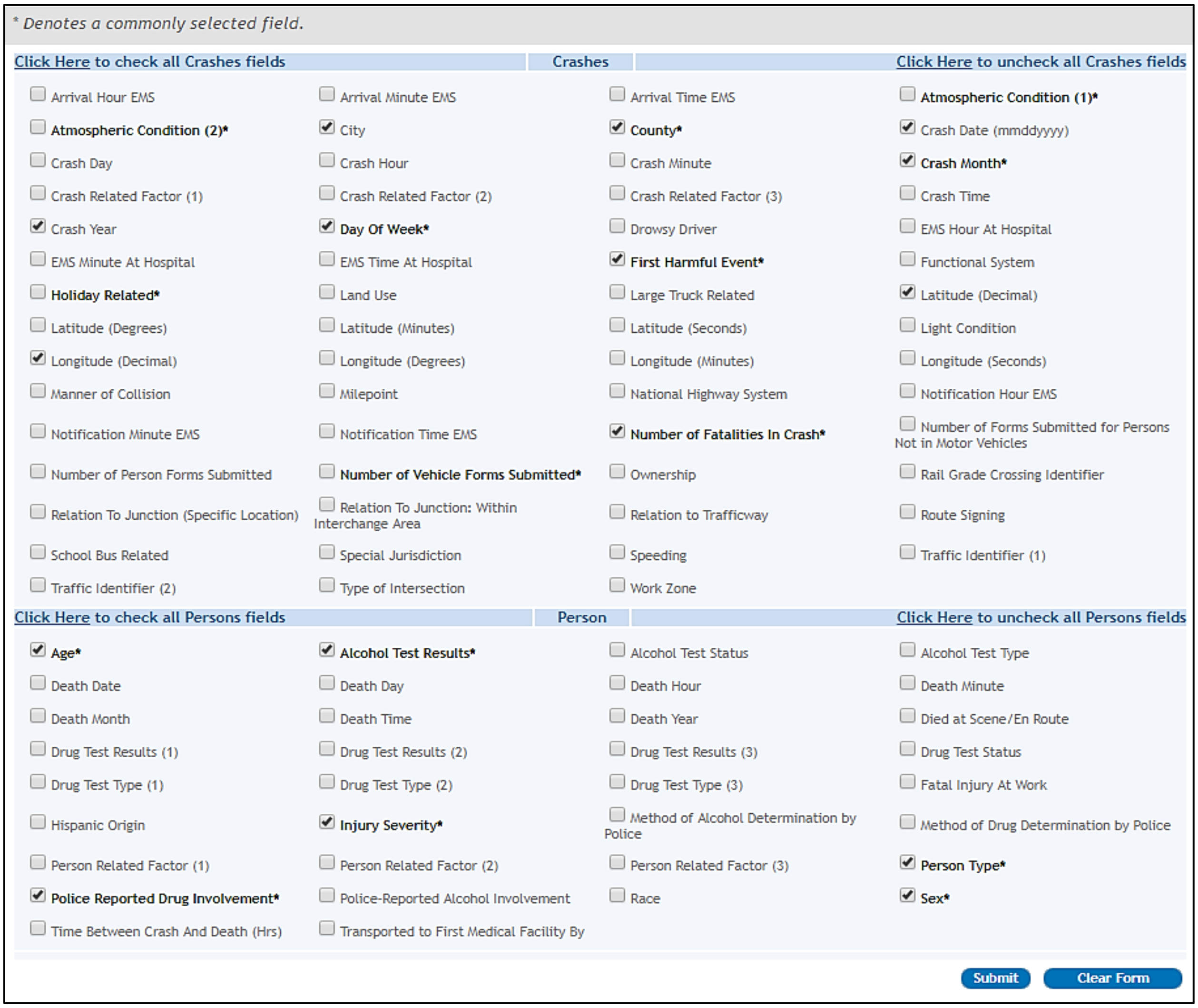
Choose Option 3 if you want to include PERSON Level information for only Occupants, and/or DRIVER Level information, PRECRASH Level information, VEHICLE Level information and /or CRASH level information

Examples uses:

- You are looking for a count of motor vehicle occupants involved in fatal crashes.
- You are looking for large trucks involved in fatal crashes where the truck jackknifed prior to its involvement.
- You are looking for crashes that involved a trafficway with a two-way continuous left turn lane where one of the vehicles was engaged in a left turn.

Step III

The required fields related to both crashes and persons were selected.



* Denotes a commonly selected field.

Click Here to check all Crashes fields	Crashes		Click Here to uncheck all Crashes fields
Arrival Hour EMS	Arrival Minute EMS	Arrival Time EMS	**Atmospheric Condition (1)***
Atmospheric Condition (2)*	City	**County***	Crash Date (mmddyyyy)
Crash Day	Crash Hour	Crash Minute	**Crash Month***
Crash Related Factor (1)	Crash Related Factor (2)	Crash Related Factor (3)	Crash Time
Crash Year	**Day Of Week***	Drowsy Driver	EMS Hour At Hospital
EMS Minute At Hospital	EMS Time At Hospital	**First Harmful Event***	Functional System
Holiday Related*	Land Use	Large Truck Related	Latitude (Decimal)
Latitude (Degrees)	Latitude (Minutes)	Latitude (Seconds)	Light Condition
Longitude (Decimal)	Longitude (Degrees)	Longitude (Minutes)	Longitude (Seconds)
Manner of Collision	Milepoint	National Highway System	Notification Hour EMS
Notification Minute EMS	Notification Time EMS	**Number of Fatalities In Crash***	Number of Forms Submitted for Persons Not in Motor Vehicles
Number of Person Forms Submitted	**Number of Vehicle Forms Submitted***	Ownership	Rail Grade Crossing Identifier
Relation To Junction (Specific Location)	Relation To Junction: Within Interchange Area	Relation to Trafficway	Route Signing
School Bus Related	Special Jurisdiction	Speeding	Traffic Identifier (1)
Traffic Identifier (2)	Type of Intersection	Work Zone	
Click Here to check all Persons fields	Person		Click Here to uncheck all Persons fields
Age*	**Alcohol Test Results***	Alcohol Test Status	Alcohol Test Type
Death Date	Death Day	Death Hour	Death Minute
Death Month	Death Time	Death Year	Died at Scene/En Route
Drug Test Results (1)	Drug Test Results (2)	Drug Test Results (3)	Drug Test Status
Drug Test Type (1)	Drug Test Type (2)	Drug Test Type (3)	Fatal Injury At Work
Hispanic Origin	**Injury Severity***	Method of Alcohol Determination by Police	Method of Drug Determination by Police
Person Related Factor (1)	Person Related Factor (2)	Person Related Factor (3)	**Person Type***
Police Reported Drug Involvement*	Police-Reported Alcohol Involvement	Race	**Sex***
Time Between Crash And Death (Hrs)	Transported to First Medical Facility By		

Submit Clear Form

Step IV

In the "States" field, Kansas was selected and in the "Age" field 15-19 was selected.

Step V

All fields were selected.

Query - Step 5: Choose Report Format Options

Submit | Clear Form

Case listing lists the cases that satisfy the query specified in previous page. One can select specific fields to be shown on the case listing table with this page. A check mark will appear to indicate that the field is selected. To de-select a field, click on the check mark again.

Total Count(s)	
Crashes:	68
Vehicles:	68
Persons:	103

Report Title:

Click Here to check all fields   Click Here to uncheck all fields

☑ State ☑ Case Number ☑ Vehicle Number ☑ Person Number

☑ County	☑ First Harmful Event	☑ Age	☑ Person Type
☑ Crash Date (mmddyyyy)	☑ Latitude (Decimal)	☑ Alcohol Test Results	☑ Police Reported Drug Involvement
☑ Crash Month	☑ Longitude (Decimal)	☑ Injury Severity	☑ Sex
☑ Day Of Week	☑ Number of Fatalities In Crash		

Search Criteria:
Year 2016
State 20
Age 15, 16, 17, 18, 19, 20

Submit | Clear Form

Step VI

The data were selected as an Excel sheet by selecting EXPORT (XLS).

Report:

Map It! | Next 100 Records

OUTPUT OPTIONS: EXPORT (TXT) EXPORT (XLS)

Obs.	State	Case Number	Vehicle Number	Person Number	County	Crash Date (mmddyyyy)	Crash Month	Day Of Week	First Harmful Event	Latitude (Decimal)	Longitude (Decimal)	Number of Fatalities In Crash	Age	Alcohol Test Results	Injury Severity	Person Type	Police Reported Drug Involvement	Sex
1	20	12	1	3	195	01152016	1	6	43	39.04324722	-99.87098333	1	18	996	1	2	8	1
2	20	17	2	1	177	01202016	1	4	12	39.12926111	-95.65452500	1	16	0	4	1	0	2
3	20	18	0	2	69	01282016	1	5	8	37.79433056	-100.23370278	2	20	0	4	5	0	1
4	20	31	1	1	173	02072016	2	1	8	37.68148611	-97.32370556	1	18	996	0	1	0	1
5	20	31	1	2	173	02072016	2	1	8	37.68148611	-97.32370556	1	19	996	0	2	8	1
6	20	41	1	1	37	02192016	2	6	42	37.39630000	-94.67244722	1	18	0	4	1	1	2
7	20	41	1	2	37	02192016	2	6	42	37.39630000	-94.67244722	1	18	996	2	2	8	2
8	20	41	1	3	37	02192016	2	6	42	37.39630000	-94.67244722	1	19	996	2	2	8	1
9	20	44	1	1	37	02282016	2	1	12	37.63887222	-94.70291944	2	17	0	4	1	0	1
10	20	44	1	2	37	02282016	2	1	12	37.63887222	-94.70291944	2	16	996	2	2	8	2
11	20	46	1	1	177	02182016	2	5	12	39.05852222	-95.72466389	1	20	-	2	1	0	1
12	20	66	2	1	173	03272016	3	1	12	37.78141667	-97.37188889	1	17	996	2	1	0	2
13	20	66	2	2	173	03272016	3	1	12	37.78141667	-97.37188889	1	19	996	2	2	8	1
14	20	68	1	1	113	04102016	4	1	12	38.28969444	-97.77773889	1	16	996	0	1	0	1
15	20	82	1	1	177	02232016	2	3	24	39.12954167	-95.85003611	1	15	996	4	1	0	1
16	20	88	1	1	123	04152016	4	6	30	39.56386667	-98.11438333	1	16	0	4	1	0	1
17	20	90	1	3	103	04282016	4	5	12	39.18675833	-94.93792500	1	17	996	3	2	8	2
18	20	91	1	1	191	05022016	5	2	12	37.14154444	-97.60167500	1	20	0	4	1	0	1

APPENDIX B

Table 29. Higher Education Schools in Kansas (IPEDS, 2019)

No.	School Name	City	Public/Private
1	Cowley County Community College	Arkansas City	Public
2	Benedictine College	Atchison	Public
3	Baker University	Baldwin City	Public
4	North Central Kansas Technical College	Beloit	Public
5	Neosho County Community College	Chanute	Public
6	Coffeyville Community College	Coffeyville	Public
7	Colby Community College	Colby	Public
8	Cloud County Community College	Concordia	Public
9	Dodge City Community College	Dodge City	Public
10	Butler Community College	El Dorado	Public
11	Flint Hills Technical College	Emporia	Public
12	Emporia State University	Emporia	Public
13	Fort Scott Community College	Fort Scott	Public
14	Garden City Community College	Garden City	Public
15	Northwest Kansas Technical College	Goodland	Public
16	Barton County Community College	Great Bend	Public
17	Barclay College	Haviland	Public
18	Fort Hays State University	Hays	Public
19	Hays Academy of Hair Design	Hays	Private
20	Hesston College	Hesston	Public
21	Highland Community College	Highland	Public
22	Tabor College	Hillsboro	Public
23	Hutchinson Community College	Hutchinson	Public
24	Sidneys Hair Dressing College	Hutchinson	Private
25	Independence Community College	Independence	Public
26	Allen County Community College	Iola	Public
27	Kansas City Kansas Community College	Kansas City	Public
28	Donnelly College	Kansas City	Public
29	Haskell Indian Nations University	Lawrence	Public
30	University of Kansas	Lawrence	Public
31	Pinnacle Career Institute-Lawrence	Lawrence	Private
32	WellSpring School of Allied Health-Lawrence	Lawrence	Private
33	University of Saint Mary	Leavenworth	Public
34	Saint Paul School of Theology	Leawood	Public
35	Brown Mackie College-Kansas City	Lenexa	Private
36	Entourage Institute of Beauty and Esthetics	Lenexa	Private

37	The Art Institutes International–Kansas City	Lenexa	Private
38	Seward County Community College	Liberal	Public
39	Bethany College	Lindsborg	Public
40	Manhattan Area Technical College	Manhattan	Public
41	Kansas State University	Manhattan	Public
42	Manhattan Christian College	Manhattan	Public
43	Bellus Academy	Manhattan	Private
44	Central Christian College of Kansas	McPherson	Public
45	McPherson College	McPherson	Public
46	Bethel College-North Newton	North Newton	Public
47	MidAmerica Nazarene University	Olathe	Public
48	Regency Beauty Institute-Olathe	Olathe	Private
49	Ottawa University-Ottawa	Ottawa	Public
50	Johnson County Community College	Overland Park	Public
51	Kansas Christian College	Overland Park	Public
52	National American University-Overland Park	Overland Park	Private
53	Cleveland University-Kansas City	Overland Park	Public
54	ITT Technical Institute-Overland Park	Overland Park	Private
55	La Baron Hairdressing Academy-Overland Park	Overland Park	Private
56	Mitsu Sato Hair Academy	Overland Park	Private
57	Ottawa University-Kansas City	Overland Park	Public
58	Paul Mitchell the School-Overland Park	Overland Park	Private
59	Wright Career College	Overland Park	Public
60	Z Hair Academy	Overland Park	Private
61	Labette Community College	Parsons	Public
62	Pittsburg State University	Pittsburg	Public
63	Pratt Community College	Pratt	Public
64	Salina Area Technical College	Salina	Public
65	Kansas Wesleyan University	Salina	Public
66	Brown Mackie College-Salina	Salina	Private
67	Hays Academy of Hair Design	Salina	Private
68	Central Baptist Theological Seminary	Shawnee	Public
69	Sterling College	Sterling	Public
70	Washburn Institute of Technology	Topeka	Public
71	Washburn University	Topeka	Public
72	Rasmussen College-Kansas	Topeka	Private
73	Bryan University	Topeka	Private
74	Regency Beauty Institute-Topeka	Topeka	Private
75	Wichita Area Technical College	Wichita	Public
76	Wichita State University	Wichita	Public

77	Vatterott College-Wichita	Wichita	Private
78	National American University-Wichita	Wichita	Private
79	National American University-Wichita West	Wichita	Private
80	Friends University	Wichita	Public
81	Newman University	Wichita	Public
82	Crave Beauty Academy	Wichita	Private
83	Eric Fisher Academy	Wichita	Private
84	Heritage College-Wichita	Wichita	Private
85	ITT Technical Institute-Wichita	Wichita	Private
86	Old Town Barber College-Wichita	Wichita	Private
87	Paul Mitchell the School-Wichita	Wichita	Private
88	Wichita Technical Institute	Wichita	Private
89	Southwestern College	Winfield	Public

Table 30. Percent of Kansas Population by Age Group (U.S. Census Bureau, 2019b)

Year\Age	2010	2011	2012	2013	2014	2015	2016	Avg.
14 and Under	21.3	21.2	21.1	21	20.8	20.6	20.5	20.8
15 to 19	7.1	7.1	7	6.9	6.9	6.9	6.9	7
20 to 24	7.2	7.2	7.4	7.5	7.6	7.6	7.5	7.4
25 to 29	6.9	6.9	6.8	6.6	6.6	6.6	6.6	6.7
30 to 34	6.3	6.5	6.6	6.7	6.7	6.7	6.6	6.6
35 to 39	6	5.9	5.9	5.9	6	6.2	6.3	6.1
40 to 44	6.1	6.1	6.1	6	5.9	5.7	5.6	5.9
45 to 49	7	6.7	6.4	6.1	5.9	5.8	5.8	6.2
50 to 54	7.2	7.2	7.1	7	6.8	6.6	6.3	6.8
55 to 59	6.4	6.5	6.6	6.7	6.7	6.7	6.7	6.6
60 to 64	5.3	5.5	5.5	5.7	5.8	5.9	6.1	5.7
65 to 69	3.8	3.9	4.2	4.3	4.5	4.8	5	4.4
70 to 74	2.9	2.9	3	3.1	3.2	3.3	3.4	3.2
75 to 79	2.4	2.4	2.4	2.4	2.4	2.4	2.5	2.4
80 to 84	2	2	2	1.9	1.9	1.9	1.9	1.9
Over 84	2.1	2.1	2.1	2.2	2.2	2.2	2.2	2.2

Table 31. Age of Teen Driver Involved in Crashes

Year\Age	15	16	17	18	19	Total
2010	710	2,580	3,059	3,323	3,225	12,897
2011	653	2,282	2,852	3,127	2,944	11,858
2012	596	2,038	2,758	3,049	2,994	11,435
2013	603	1,841	2,789	3,096	3,004	11,333
2014	670	1,920	2,740	3,038	2,990	11,358
2015	668	1,998	2,824	3,002	3,053	11,545
2016	734	2,084	3,049	3,211	3,060	12,138
Average	662	2,106	2,867	3,121	3,039	11,795
Percent	5.61	17.86	24.31	26.46	25.76	

Table 32. Number of Teen Drivers Involved in Crashes Related to Driving Under the Influence

Year	Alcohol-Related			Not Alcohol-Related			DUI	Sober	Total
	Drug-Related	Not Drug-Related	Total	Drug-Related	Not Drug-Related	Total			
2010	19	249	268	18	12,611	12,629	286	12,611	12,897
2011	22	223	245	26	11,587	11,613	271	11,587	11,858
2012	21	260	281	17	11,137	11,154	298	11,137	11,435
2013	13	166	179	37	11,117	11,154	216	11,117	11,333
2014	11	140	151	32	11,175	11,207	183	11,175	11,358
2015	13	161	174	22	11,349	11,371	196	11,349	11,545
2016	14	143	157	34	11,947	11,981	191	11,947	12,138
Total	113	1,342	1,455	186	80,923	81,109	1,641	80,923	82,564

Table 33. Teen Driver Fatalities Involved in Crashes Related to Driving Under the Influence

Year	Alcohol-Related			Not Alcohol-Related			DUI		Sober	Total
	Drug-Related	Not Drug-Related	Total	Drug-Related	Not Drug-Related	Total	(#)	(%)		
2010	0	6	6	0	21	21	6	22.22%	21	27
2011	3	2	5	0	15	15	5	25.00%	15	20
2012	3	5	8	0	13	13	8	38.10%	13	21
2013	2	1	3	1	12	13	4	25.00%	12	16
2014	1	3	4	5	14	19	9	39.13%	14	23
2015	1	2	3	1	17	18	4	19.05%	17	21
2016	1	2	3	2	17	19	5	22.73%	17	22
Total	11	21	32	9	109	118	41	27.33%	109	152

Table 34. Injury Severity of Teen Drivers by Vehicle Body Type

Body Type	K		A		B		C		O	
	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)
PC	90	60.00	418	58.22	2,935	61.49	3,872	67.53	45,459	64.77
Motorcycle, Moped, Scooter	7	4.67	54	7.52	170	3.56	83	1.45	98	0.14
Pickup Track	29	19.33	126	17.55	806	16.89	820	14.30	11,510	16.40
SUV	18	12.00	91	12.67	739	15.48	813	14.18	11,127	15.85
Van	2	1.33	19	2.65	72	1.51	119	2.08	1,764	2.51
Others	4	2.67	10	1.39	51	1.07	27	0.47	223	0.32
Total	150		718		4,773		5,734		70,181	

Table 35. Model Year of Vehicles Involved in Crashes That Driven by Teen Drivers

Model Year	2010	2011	2012	2013	2014	2015	2016	Total
1900	1[1]							1
1909			1					1
1937						1		1
1946					1			1
1948					1			1
1949					1			1
1955	1	1		1				3
1960	1							1
1963	1	1			4			6
1964			1	1	1	1		4
1965		2		1	2	1		6
1966	2		4	1	2	1	1	11
1967	2	1	3		6	1	2	15
1968	5	2	1		1		1	10
1969	2	1	2	1	4	3	1	14
1970	2	1	2	6	4	3	4	22
1971	5	1	2	4	4	2		18
1972	13	1	8	5	5	1	3	36
1973	4	2	6	2	4	3	2	23
1974	6	1	2	2	2	2	2	17
1975	7	3	4	2	6	3	2	27
1976	7	6	8	3	4	3	5	36
1977	9	8	11	5	8	2	4	47
1978	23	12	12	6	4	3	5	65
1979	15	8	17	10	8	9	8	75
1980	14	12	8	7	4	11	8	64
1981	13	10	9	11	5	13		61
1982	25	19	10	13	13	13	7	100
1983	16	16	17	18	12	8	11	98
1984	36	32	20	24	11	10	10	143

[1] It is possible the model year numerals of these vehicles were transposed or inputted incorrectly. This perspective is probably true for other reported model year such as 1955 and 1960 or it might even go further. Since it is unknown exactly where that breakpoint is, for simplicity, they all have been retained in the analysis.

1985	46	32	28	24	26	16	16	188
1986	68	47	40	42	37	17	18	269
1987	68	54	37	34	29	26	18	266
1988	102	59	67	41	38	33	26	366
1989	136	97	81	66	57	37	44	518
1990	176	130	113	101	71	65	39	695
1991	203	161	140	110	98	88	70	870
1992	269	241	189	151	112	128	99	1,189
1993	363	265	217	218	181	136	140	1,520
1994	502	390	331	301	249	206	182	2,161
1995	605	459	428	314	309	271	218	2,604
1996	658	552	450	401	343	283	282	2,969
1997	825	699	606	576	475	451	398	4,030
1998	901	783	663	603	600	512	436	4,498
1999	996	912	854	766	688	650	617	5,483
2000	1,031	932	944	833	799	808	703	6,050
2001	954	920	871	932	801	834	754	6,066
2002	909	954	861	866	842	849	798	6,079
2003	761	753	709	740	821	820	824	5,428
2004	737	686	726	768	783	808	867	5,375
2005	600	627	668	670	784	753	865	4,967
2006	491	497	522	613	681	692	813	4,309
2007	464	437	456	514	535	659	732	3,797
2008	361	372	390	415	490	548	673	3,249
2009	223	233	239	246	241	297	347	1,826
2010	124	193	211	264	276	284	354	1,706
2011	14	117	184	210	219	282	317	1,343
2012		15	174	209	239	289	382	1,308
2013			18	143	258	258	349	1,026
2014				15	133	217	284	649
2015					7	107	242	356
2016						8	117	125
2017							11	11
Total	12,797	11,757	11,365	11,309	11,339	11,526	12,111	82,204
>=5 yrs	86.90	88.37	89.30	90.39	90.02	89.93	88.56	

Table 36. Body Type of Vehicle Involved in Crashes by Age and Gender of Teen Drivers

Age	Gender	PC	Motorcycle, Moped, Scooter	Pickup Track	SUV	Van	Others	Total
15	Female	1,272	4	292	498	71	8	2,145
	Male	1,096	11	817	457	82	17	2,480
16	Female	4,755	9	583	1,459	210	10	7,026
	Male	4,003	26	2,161	1,284	192	30	7,696
17	Female	6,934	7	688	1,714	231	7	9,581
	Male	6,075	56	2,469	1,578	247	42	10,467
18	Female	7,746	17	643	1,539	249	14	10,208
	Male	6,949	92	2,683	1,546	278	66	11,614
19	Female	7,602	19	546	1,456	190	12	9,825
	Male	6,917	177	2,567	1,398	245	115	11,419
Total Female	(No.)	28,309	56	2,752	6,666	951	51	38,785
	(%)	53.06	13.4	20.46	51.56	47.67	15.89	
Total Male	(No.)	25,040	362	10,697	6,263	1,044	270	43,676
	(%)	46.94	86.6	79.54	48.44	52.33	84.11	

Table 37. The Severity of Crashes on Local Roads and State Highways

Road System	Local Roadway				State Highway System				
Severity	Fatal	Injury	PDO	Total	Fatal	Injury	PDO	Total	Total
2010	34	2198	6425	8657	20	777	2311	3108	11765
2011	19	2039	6101	8159	18	763	2038	2819	10978
2012	33	2095	5903	8031	16	669	1871	2556	10587
2013	23	1918	5948	7889	16	662	1905	2583	10472
2014	19	1884	5877	7780	22	689	2017	2728	10508
2015	21	1915	5972	7908	17	686	2098	2801	10709
2016	22	1981	6305	8308	20	651	2193	2864	11172
Total	171	14030	42531	56732	129	4897	14433	19459	76191

Table 38. Crash Severity by Road Function Classes

Road Function Class	Fatal		Injury		PDO		Total	
	(No.)	(%)	(No.)	(%)	(No.)	(%)	(No.)	(%)
Interstate	-	-	7	0.05	16	0.04	23	0.04
Local	80	46.78	4,546	32.4	14,929	35.1	19,555	34.47
Major Collector	48	28.07	2,217	15.8	6,314	14.85	8,579	15.12
Min. Arterial	20	11.7	3,416	24.35	10,310	24.24	13,746	24.23
Min Collector	11	6.43	374	2.67	751	1.77	1,136	2
Pr. Arterial	12	7.02	3,470	24.73	10,211	24.01	13,693	24.14
Total	171		14,030		42,531		56,732	

Table 39. Crashes Involving Teen Drivers and Light Conditions

Light Conditions	Fatal	Injury	PDO	Total
Dark (No Street Lights)	89	2,103	6,183	8,375
Dark (Street Lights On)	32	2,794	8,643	11,469
Dark	121	4,897	14,826	19,844
Daylight	160	13,133	39,290	52,583
Dusk	11	537	1,570	2,118
Dawn	7	335	1,160	1,502
Dusk and Dawn (DD)	18	872	2,730	3,620
N_K [1](crash/hr.)	10.5	425.8	1,289.2	1,725.6
N_D [2](crash/hr.)	13.9	1,142.0	3,416.5	4,572.4
Dark Difference[3] (%)	41.6	51.2	52.8	52.3
Daylight Difference[3] (%)	22.7	-30.9	-25.2	-26.3
Daylight and Dark Difference[4] (%)	32.1	10.1	13.8	13.0

[1] We accordingly introduce a measure, N, that effectively normalizes for the prevalence of the light conditions. This is defined for dark condition as: N_K=K/11.5 where K is number of crashes during dark condition and 11.5 represent number of hours of dark condition.

[2] Similarly for Daylight condition, N_D=K/11.5 where D is number of crashes during daylight condition and 11.5 represent number of hours of daylight condition. Since duration of dawn and dusk is one hour, normalizing was not needed.

[3] Dark Difference is difference between DD and N_D in percent. Similarly for Daylight Difference.

[4] Daylight and Dark Difference is the average of Dark Difference and Daylight Difference.

Table 40. Crash Severity by Day of the Week

Day\Crash Severity	FATAL	INJURY	PDO
Saturday	51	2,665	7,563
Sunday	67	2,215	5,878
Monday	40	2,552	8,022
Tuesday	31	2,660	8,095
Wednesday	38	2,734	8,569
Thursday	24	2,763	8,522
Friday	45	3,310	10,210

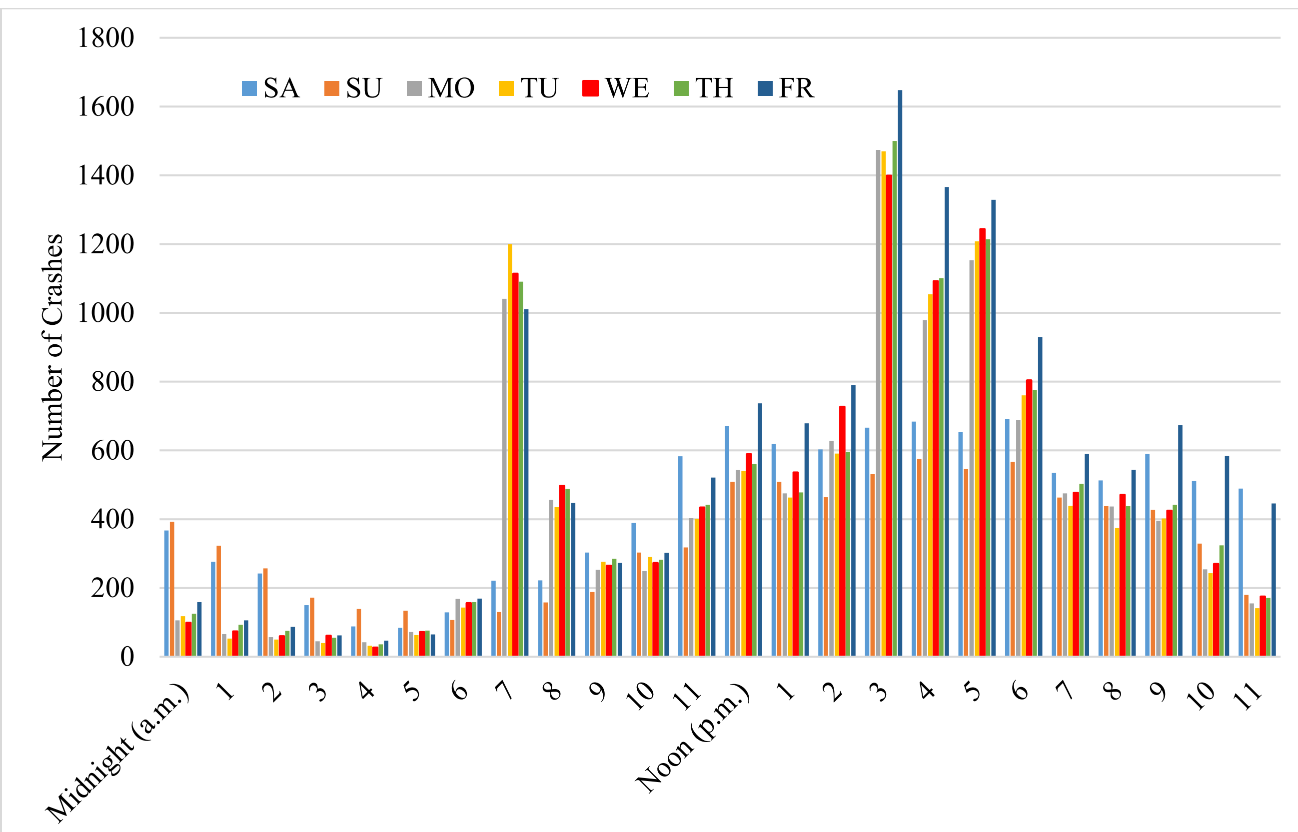


Figure 59. Number of Crashes by Day of the Week and Time of the Day

Table 41. Severity of Crashes by Gender and Year

Severity	Gender	2010	2011	2012	2013	2014	2015	2016	Total	Rate (%)
K	F	10	6	6	6	6	8	8	50	33.33
	M	17	14	15	10	17	13	14	100	66.67
	Sum	27	20	21	16	23	21	22	150	
A	F	68	50	50	49	32	32	35	316	44.01
	M	82	69	54	49	50	51	46	401	55.85
	Sum	151	119	104	98	82	83	81	718	
B	F	380	366	344	325	337	315	275	2,342	49.07
	M	369	353	368	324	298	364	351	2,427	50.85
	Sum	752	720	712	649	635	679	626	4,773	
C	F	552	464	476	443	444	469	489	3,337	58.15
	M	354	340	365	347	343	324	326	2,399	41.80
	Sum	909	804	841	790	787	793	815	5,739	
O	F	5,007	4,511	4,433	4,459	4,558	4,565	4,829	32,362	46.10
	M	5,775	5,455	5,197	5,217	5,191	5,302	5,662	37,799	53.84
	Sum	10,794	9,975	9,634	9,680	9,752	9,875	10,494	70,204	

Table 42. The Severity of CWOVs

CWOV	Fatal		Injury		PDO		Total	
	No.	(%)	No.	(%)	No.	(%)	No.	(%)
Rear End	10	7.04	5,401	43.03	17,639	44.76	23,050	44.24
Angle	87	61.27	5,791	46.13	15,108	38.34	20,986	40.28
Sideswipe	11	7.75	585	4.66	4,394	11.15	4,990	9.58
Backed Into	-	0.00	42	0.33	1,195	3.03	1,237	2.37
Head On	34	23.94	729	5.81	998	2.53	1,761	3.38
Other	-	0.00	5	0.04	72	0.18	77	0.15
Total	142	100	12,553	100	39,406	100	52,101	100

APPENDIX C

The utilized crashes involving teen drivers' data from the previous steps were in Excel Workbook formats. To convert those crash data points to shapefiles through ArcGIS software, the Excel Workbooks were saved as Comma-separated Values (CSV) (Comma delimited) files. From the Add Data in the Map tab in the ArcGIS Pro software ribbon, the CSV files were selected using the XY Point Data tab. The CSV file was selected as an Input Table and a desired name and location were selected for saving in the Output Feature Class field, as shown in Figure 60. The longitude column of the CSV file was selected as the X Field and the Latitude column as the Y Field. In the Coordinate System field, the NAD83 Kansas LCC coordinate system was used for spatial analysis at the state level and the NAD_1983_UTM_Zone_14N or 15N was used for districts and counties level, based on their location.

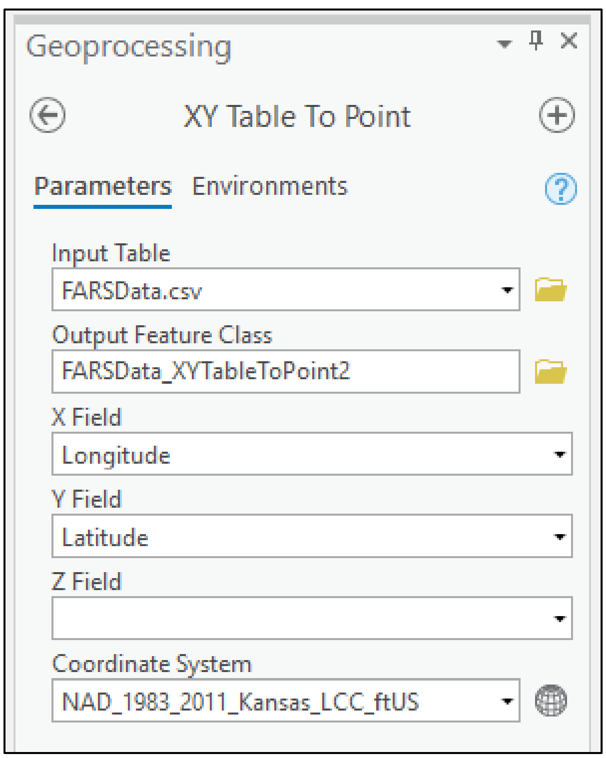


Figure 60. Converting Crash Points to a Shape File

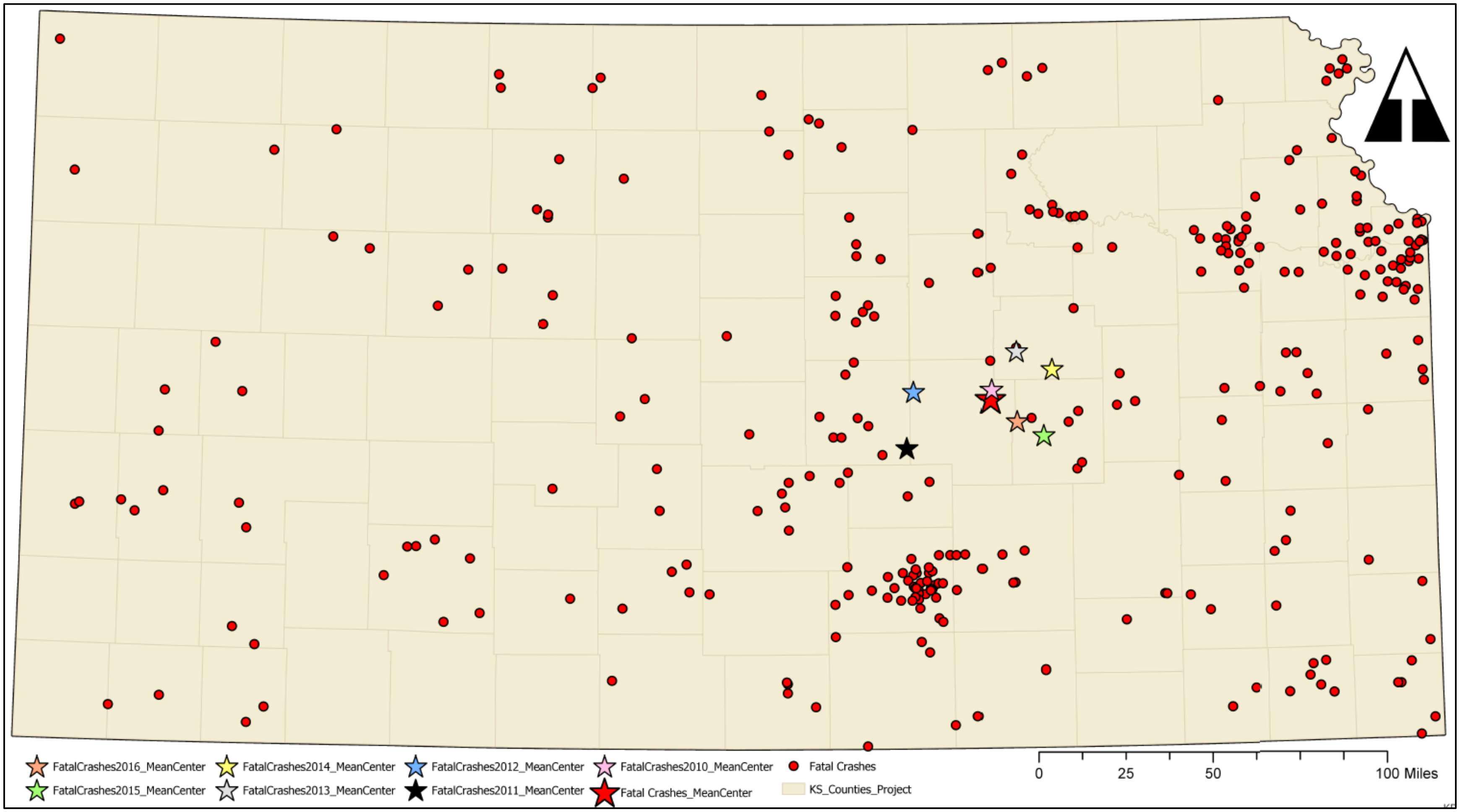


Figure 61. Mean Centers for Fatal Crashes Involving Teen Drivers by Year

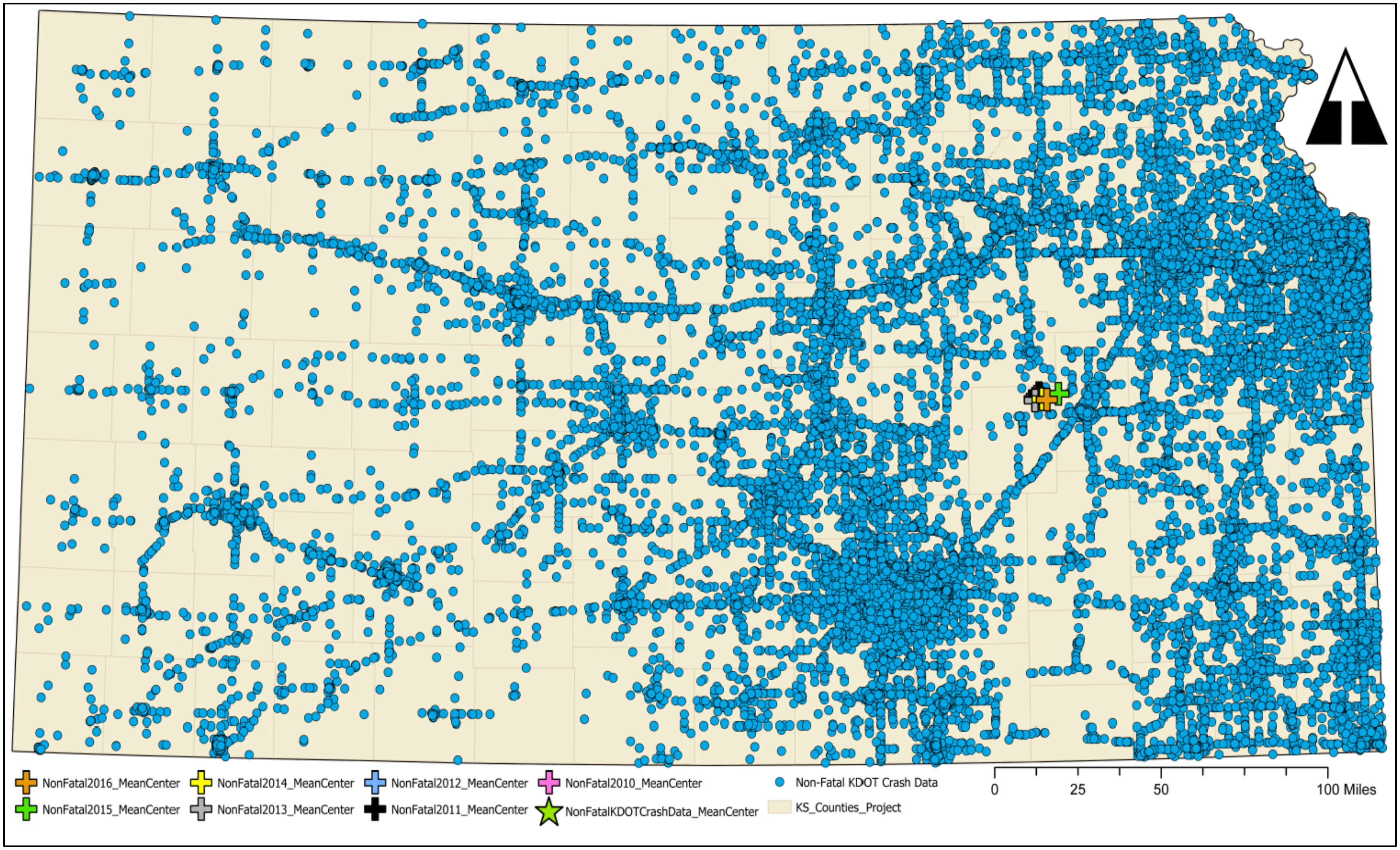


Figure 62. Mean Centers for Non-Fatal Crashes Involving Teen Drivers by Year

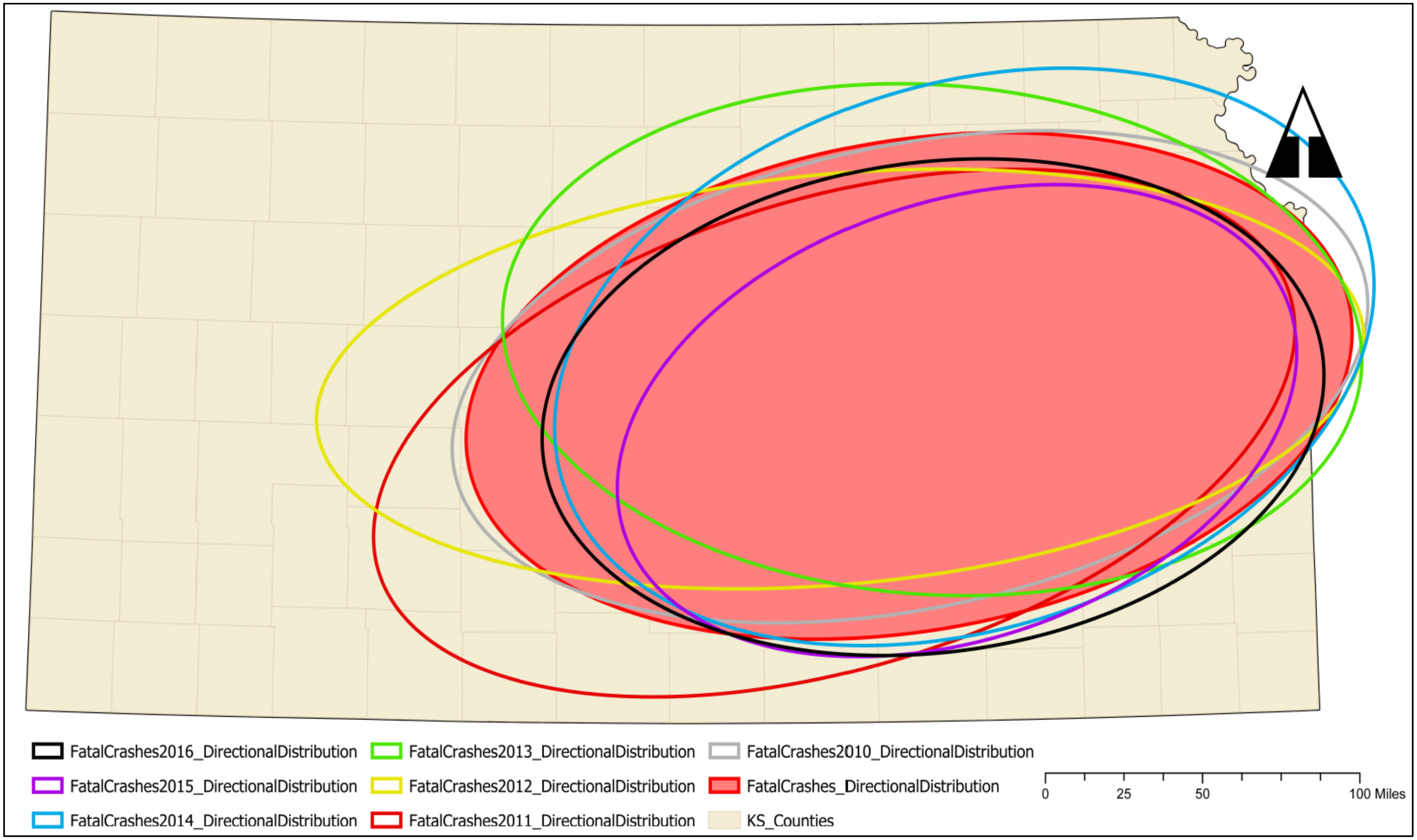


Figure 63. Directional Distribution for Fatal Crashes Involving Teen Drivers by Year

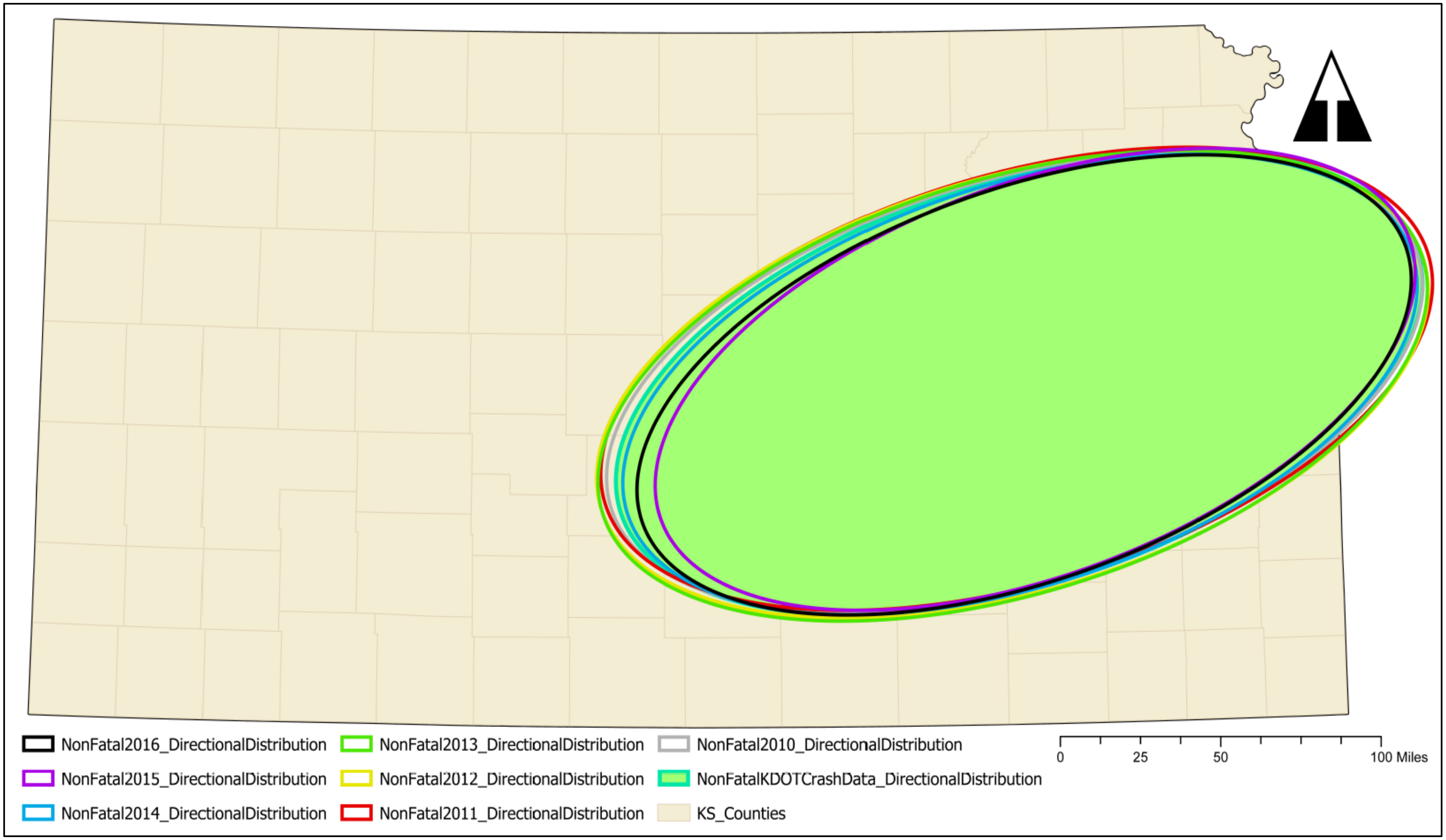
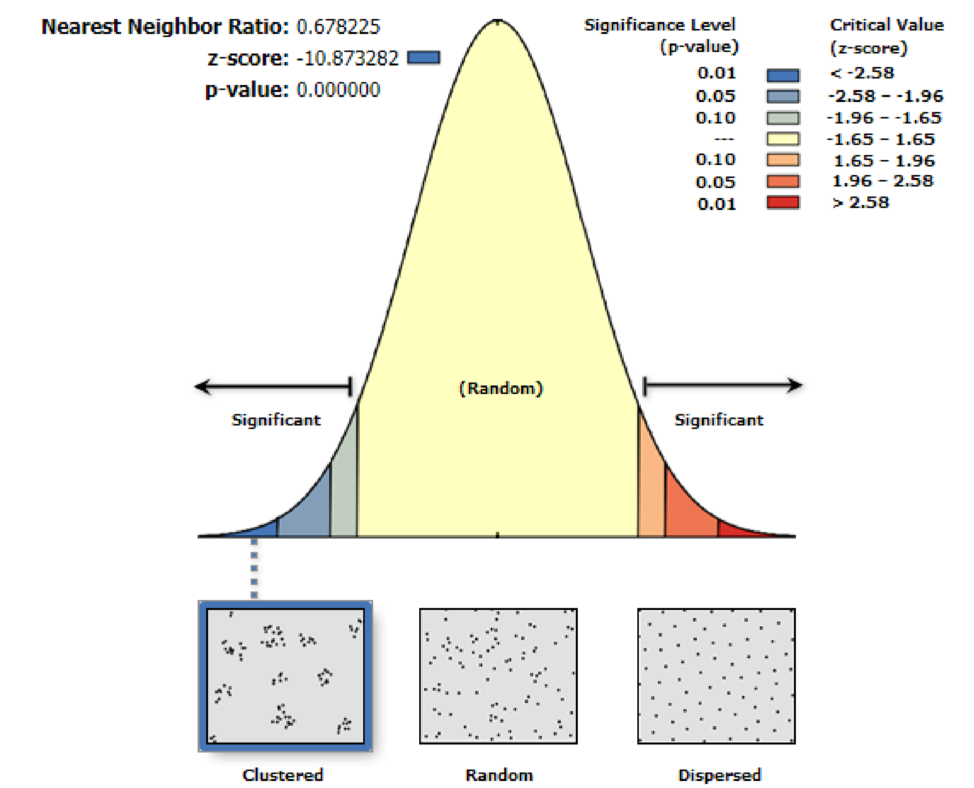


Figure 64. Directional Distribution for Non-fatal Crashes Involving Teen Drivers by Year



Given the z-score of -10.873282, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Average Nearest Neighbor Summary

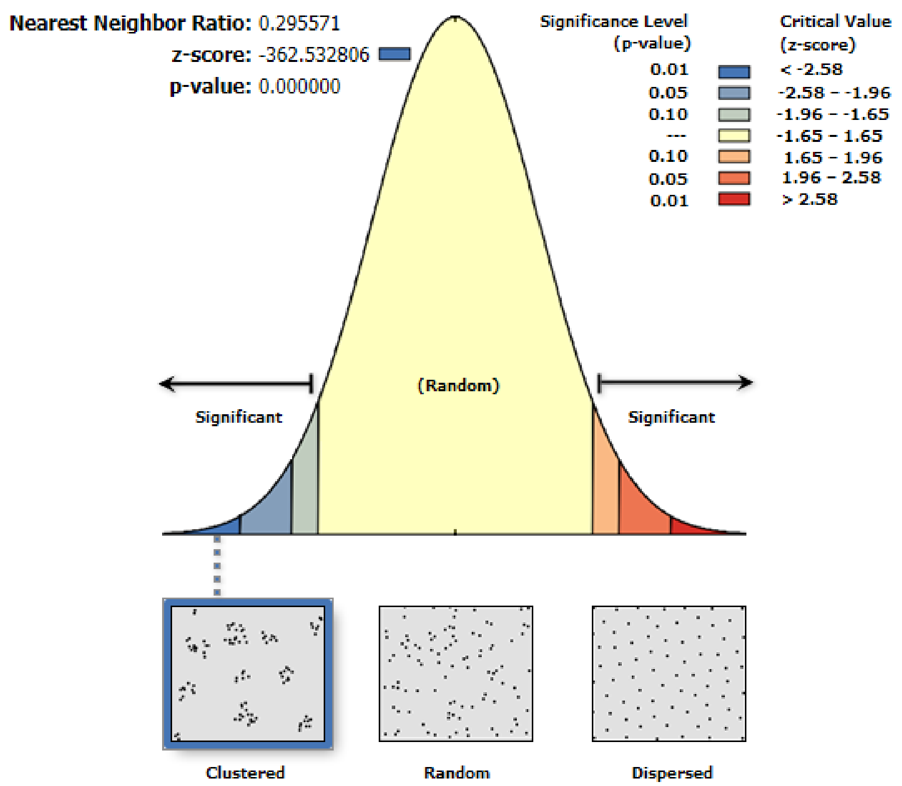
Observed Mean Distance:	29076.2964 US_Feet
Expected Mean Distance:	42871.1741 US_Feet
Nearest Neighbor Ratio:	0.678225
z-score:	-10.873282
p-value:	0.000000

Dataset Information

Input Feature Class:	Fatal Crashes
Distance Method:	EUCLIDEAN
Study Area:	2293746084000.000000
Selection Set:	False

Figure 65. The Average Nearest Neighbor Report on Fatal Crashes

Average Nearest Neighbor Summary



Given the z-score of -362.532806, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Average Nearest Neighbor Summary

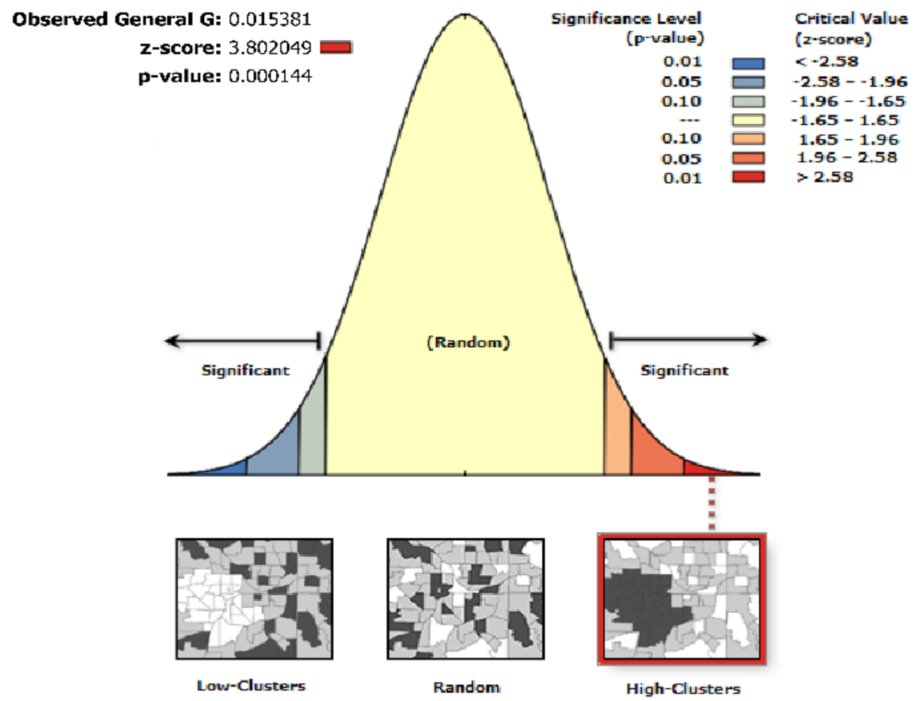
Observed Mean Distance:	832.0040 US_Feet
Expected Mean Distance:	2814.9027 US_Feet
Nearest Neighbor Ratio:	0.295571
z-score:	-362.532806
p-value:	0.000000

Dataset Information

Input Feature Class:	Non-Fatal KDOT Crash Data
Distance Method:	EUCLIDEAN
Study Area:	2293746084000.000000
Selection Set:	False

Figure 66. The Average Nearest Neighbor Report on Non-fatal Crashes

High-Low Clustering Report



Given the z-score of 3.8020489738435996, there is a less than 1% likelihood that this high-clustered pattern could be the result of random chance.

General G Summary

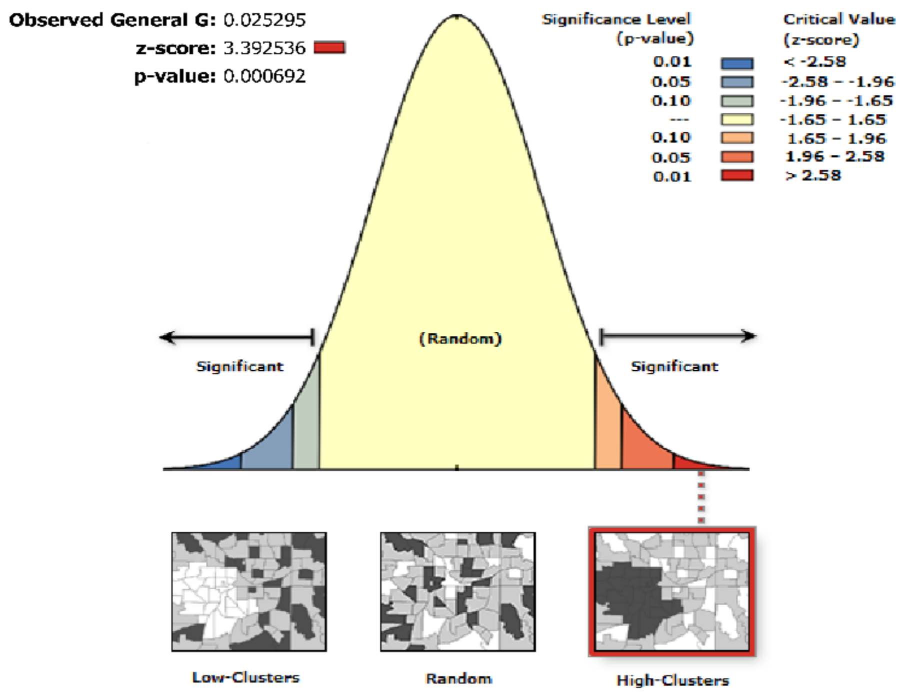
Observed General G:	0.015381
Expected General G:	0.009615
Variance:	0.000002
z-score:	3.802049
p-value:	0.000144

Dataset Information

Input Feature Class:	FatalCrashes_SpatialJoin
Input Field:	JOIN_COUNT
Conceptualization:	CONTIGUITY_EDGES_CORNERS
Distance Method:	EUCLIDEAN
Row Standardization:	True
Distance Threshold:	None
Weights Matrix File:	None
Selection Set:	False

Figure 67. High/Low Clustering (Getis-Ord General G) Report on Fatal Crashes

High-Low Clustering Report



Given the z-score of 3.392535560766044, there is a less than 1% likelihood that this high-clustered pattern could be the result of random chance.

General G Summary

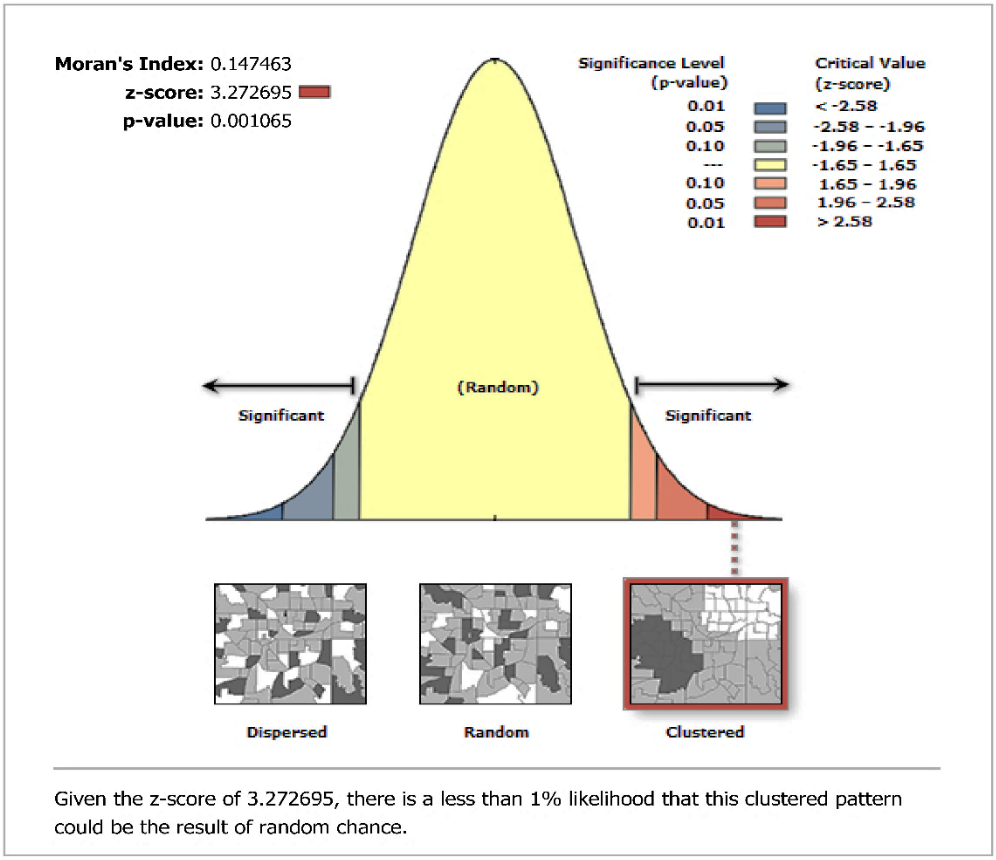
Observed General G:	0.025295
Expected General G:	0.009615
Variance:	0.000021
z-score:	3.392536
p-value:	0.000692

Dataset Information

Input Feature Class:	Non-fatalCrashes_SpatialJoin
Input Field:	JOIN_COUNT
Conceptualization:	CONTIGUITY_EDGES_CORNERS
Distance Method:	EUCLIDEAN
Row Standardization:	True
Distance Threshold:	None
Weights Matrix File:	None
Selection Set:	False

Figure 68. High/Low Clustering (Getis-Ord General G) Report on Fatal Crashes

Spatial Autocorrelation Report



Global Moran's I Summary

Moran's Index:	0.147463
Expected Index:	-0.009615
Variance:	0.002304
z-score:	3.272695
p-value:	0.001065

Dataset Information

Input Feature Class:	Non-fatalCrashes_SpatialJoin
Input Field:	JOIN_COUNT
Conceptualization:	CONTIGUITY_EDGES_CORNERS
Distance Method:	EUCLIDEAN
Row Standardization:	True
Distance Threshold:	None
Weights Matrix File:	None
Selection Set:	False

Figure 69. Spatial Autocorrelation Report on Non-Fatal Crashes

Optimized Hot Spot Analysis (Spatial Statistics Tools)

Completed Today at 2:26:00 PM

Parameters

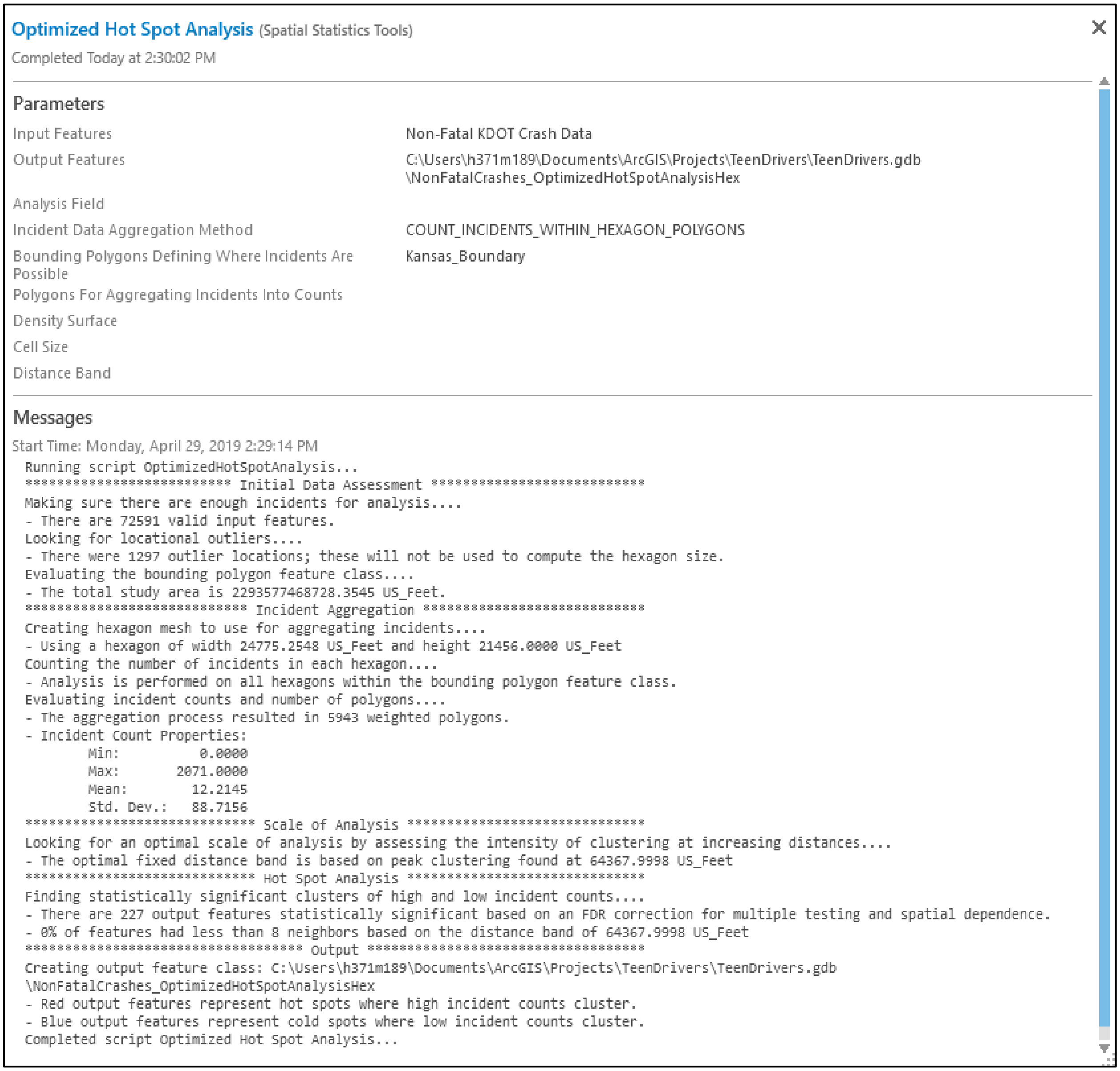
Input Features	Fatal Crashes
Output Features	C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb \FatalCrashes_OptimizedHotSpotAnalysisHex
Analysis Field	
Incident Data Aggregation Method	COUNT_INCIDENTS_WITHIN_HEXAGON_POLYGONS
Bounding Polygons Defining Where Incidents Are Possible	Kansas_Boundary
Polygons For Aggregating Incidents Into Counts	
Density Surface	
Cell Size	
Distance Band	

Messages

Start Time: Monday, April 29, 2019 2:25:35 PM

```
Running script OptimizedHotSpotAnalysis...
*************************** Initial Data Assessment ****************************
Making sure there are enough incidents for analysis....
- There are 312 valid input features.
Looking for locational outliers....
- There were 6 outlier locations; these will not be used to compute the hexagon size.
Evaluating the bounding polygon feature class....
- The total study area is 2293577468728.3545 US_Feet.
***************************** Incident Aggregation ******************************
Creating hexagon mesh to use for aggregating incidents....
- Using a hexagon of width 65013.1044 US_Feet and height 56303.0000 US_Feet
Counting the number of incidents in each hexagon....
- Analysis is performed on all hexagons within the bounding polygon feature class.
Evaluating incident counts and number of polygons....
- The aggregation process resulted in 907 weighted polygons.
- Incident Count Properties:
        Min:        0.0000
        Max:       16.0000
        Mean:       0.3440
        Std. Dev.:  1.0324
****************************** Scale of Analysis ******************************
Looking for an optimal scale of analysis by assessing the intensity of clustering at increasing distances....
- The optimal fixed distance band is based on peak clustering found at 112606.9999 US_Feet
****************************** Hot Spot Analysis ******************************
Finding statistically significant clusters of high and low incident counts....
- There are 61 output features statistically significant based on an FDR correction for multiple testing and spatial dependence.
- .2% of features had less than 8 neighbors based on the distance band of 112606.9999 US_Feet
************************************ Output ************************************
Creating output feature class: C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb
\FatalCrashes_OptimizedHotSpotAnalysisHex
- Red output features represent hot spots where high incident counts cluster.
- Blue output features represent cold spots where low incident counts cluster.
Completed script Optimized Hot Spot Analysis...
```

Figure 70. The Optimized Hot Spot Analysis Report for Fatal Crashes



Optimized Hot Spot Analysis (Spatial Statistics Tools)

Completed Today at 2:30:02 PM

Parameters

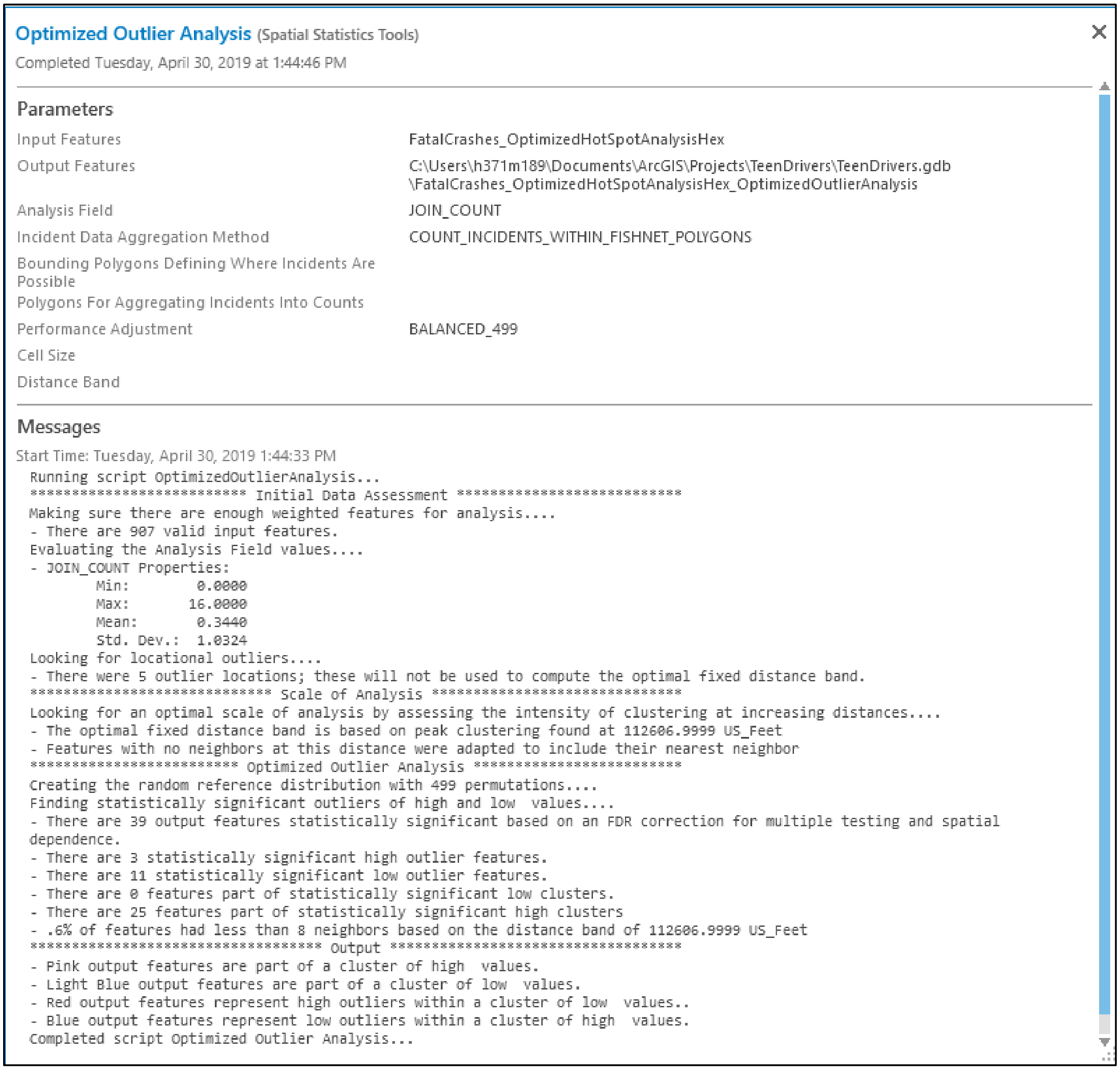
Input Features	Non-Fatal KDOT Crash Data
Output Features	C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb \NonFatalCrashes_OptimizedHotSpotAnalysisHex
Analysis Field	
Incident Data Aggregation Method	COUNT_INCIDENTS_WITHIN_HEXAGON_POLYGONS
Bounding Polygons Defining Where Incidents Are Possible	Kansas_Boundary
Polygons For Aggregating Incidents Into Counts	
Density Surface	
Cell Size	
Distance Band	

Messages

Start Time: Monday, April 29, 2019 2:29:14 PM

```
Running script OptimizedHotSpotAnalysis...
************************** Initial Data Assessment ***************************
Making sure there are enough incidents for analysis....
- There are 72591 valid input features.
Looking for locational outliers....
- There were 1297 outlier locations; these will not be used to compute the hexagon size.
Evaluating the bounding polygon feature class....
- The total study area is 2293577468728.3545 US_Feet.
**************************** Incident Aggregation ****************************
Creating hexagon mesh to use for aggregating incidents....
- Using a hexagon of width 24775.2548 US_Feet and height 21456.0000 US_Feet
Counting the number of incidents in each hexagon....
- Analysis is performed on all hexagons within the bounding polygon feature class.
Evaluating incident counts and number of polygons....
- The aggregation process resulted in 5943 weighted polygons.
- Incident Count Properties:
        Min:          0.0000
        Max:       2071.0000
        Mean:        12.2145
        Std. Dev.:   88.7156
***************************** Scale of Analysis ******************************
Looking for an optimal scale of analysis by assessing the intensity of clustering at increasing distances....
- The optimal fixed distance band is based on peak clustering found at 64367.9998 US_Feet
***************************** Hot Spot Analysis ******************************
Finding statistically significant clusters of high and low incident counts....
- There are 227 output features statistically significant based on an FDR correction for multiple testing and spatial dependence.
- 0% of features had less than 8 neighbors based on the distance band of 64367.9998 US_Feet
*********************************** Output ***********************************
Creating output feature class: C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb
\NonFatalCrashes_OptimizedHotSpotAnalysisHex
- Red output features represent hot spots where high incident counts cluster.
- Blue output features represent cold spots where low incident counts cluster.
Completed script Optimized Hot Spot Analysis...
```

Figure 71. The Optimized Hot Spot Analysis Report for Non-Fatal Crashes



Optimized Outlier Analysis (Spatial Statistics Tools)

Completed Tuesday, April 30, 2019 at 1:44:46 PM

Parameters

Input Features	FatalCrashes_OptimizedHotSpotAnalysisHex
Output Features	C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb \FatalCrashes_OptimizedHotSpotAnalysisHex_OptimizedOutlierAnalysis
Analysis Field	JOIN_COUNT
Incident Data Aggregation Method	COUNT_INCIDENTS_WITHIN_FISHNET_POLYGONS
Bounding Polygons Defining Where Incidents Are Possible	
Polygons For Aggregating Incidents Into Counts	
Performance Adjustment	BALANCED_499
Cell Size	
Distance Band	

Messages

Start Time: Tuesday, April 30, 2019 1:44:33 PM

```
Running script OptimizedOutlierAnalysis...
************************** Initial Data Assessment ***************************
Making sure there are enough weighted features for analysis....
- There are 907 valid input features.
Evaluating the Analysis Field values....
- JOIN_COUNT Properties:
        Min:        0.0000
        Max:       16.0000
        Mean:       0.3440
        Std. Dev.:  1.0324
Looking for locational outliers....
- There were 5 outlier locations; these will not be used to compute the optimal fixed distance band.
***************************** Scale of Analysis ******************************
Looking for an optimal scale of analysis by assessing the intensity of clustering at increasing distances....
- The optimal fixed distance band is based on peak clustering found at 112606.9999 US_Feet
- Features with no neighbors at this distance were adapted to include their nearest neighbor
************************* Optimized Outlier Analysis *************************
Creating the random reference distribution with 499 permutations....
Finding statistically significant outliers of high and low  values....
- There are 39 output features statistically significant based on an FDR correction for multiple testing and spatial
dependence.
- There are 3 statistically significant high outlier features.
- There are 11 statistically significant low outlier features.
- There are 0 features part of statistically significant low clusters.
- There are 25 features part of statistically significant high clusters
- .6% of features had less than 8 neighbors based on the distance band of 112606.9999 US_Feet
*********************************** Output ***********************************
- Pink output features are part of a cluster of high  values.
- Light Blue output features are part of a cluster of low  values.
- Red output features represent high outliers within a cluster of low  values..
- Blue output features represent low outliers within a cluster of high  values.
Completed script Optimized Outlier Analysis...
```

Figure 72. The Optimized Cluster and Outlier Analysis Report for Fatal Crashes

Optimized Outlier Analysis (Spatial Statistics Tools)

Completed Today at 9:00:12 AM

Parameters

Input Features	NonFatalCrashes_OptimizedHotSpotAnalysisHex
Output Features	C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb \NonFatalCrashes_OptimizedHotSpotAnalysisHex_OptimizedOutlierAnalysis
Analysis Field	JOIN_COUNT
Incident Data Aggregation Method	COUNT_INCIDENTS_WITHIN_FISHNET_POLYGONS
Bounding Polygons Defining Where Incidents Are Possible	
Polygons For Aggregating Incidents Into Counts	
Performance Adjustment	BALANCED_499
Cell Size	
Distance Band	

Messages

Start Time: Wednesday, May 1, 2019 8:59:45 AM

```
Running script OptimizedOutlierAnalysis...
************************** Initial Data Assessment ***************************
Making sure there are enough weighted features for analysis....
- There are 5943 valid input features.
Evaluating the Analysis Field values....
- JOIN_COUNT Properties:
        Min:          0.0000
        Max:       2071.0000
        Mean:        12.2145
        Std. Dev.:   88.7156
Looking for locational outliers....
- There were 93 outlier locations; these will not be used to compute the optimal fixed distance band.
****************************** Scale of Analysis ******************************
Looking for an optimal scale of analysis by assessing the intensity of clustering at increasing distances....
- The optimal fixed distance band is based on peak clustering found at 64367.9998 US_Feet
- Features with no neighbors at this distance were adapted to include their nearest neighbor
************************* Optimized Outlier Analysis *************************
Creating the random reference distribution with 499 permutations....
Finding statistically significant outliers of high and low  values....
- There are 2274 output features statistically significant based on an FDR correction for multiple testing and spatial dependence.
- There are 26 statistically significant high outlier features.
- There are 52 statistically significant low outlier features.
- There are 2047 features part of statistically significant low clusters.
- There are 149 features part of statistically significant high clusters
- 0% of features had less than 8 neighbors based on the distance band of 64367.9998 US_Feet
************************************ Output ***********************************
- Pink output features are part of a cluster of high  values.
- Light Blue output features are part of a cluster of low  values.
- Red output features represent high outliers within a cluster of low  values..
- Blue output features represent low outliers within a cluster of high  values.
Completed script Optimized Outlier Analysis...
```

Figure 73. The Optimized Cluster and Outlier Analysis Report for Non-Fatal Crashes

Figure 74. The Resultant Standard Deviational Ellipse of the USD 259

Optimized Hot Spot Analysis (Spatial Statistics Tools)

Completed Today at 11:49:29 AM

Parameters

Input Features	Fatal_Injury_inUSD259
Output Features	C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb \Fatal_Injury_inUSD259_OptimizedHotSpotAnalysis
Analysis Field	
Incident Data Aggregation Method	COUNT_INCIDENTS_WITHIN_HEXAGON_POLYGONS
Bounding Polygons Defining Where Incidents Are Possible	USD_259
Polygons For Aggregating Incidents Into Counts	
Density Surface	
Cell Size	
Distance Band	

Messages

Start Time: Sunday, May 5, 2019 11:48:49 AM

```
Running script OptimizedHotSpotAnalysis...
************************** Initial Data Assessment ***************************
Making sure there are enough incidents for analysis....
- There are 3019 valid input features.
Looking for locational outliers....
- There were 43 outlier locations; these will not be used to compute the hexagon size.
Evaluating the bounding polygon feature class....
- The total study area is 4245635781.2126 US_Feet.
**************************** Incident Aggregation ****************************
Creating hexagon mesh to use for aggregating incidents....
- Using a hexagon of width 1096.9655 US_Feet and height 950.0000 US_Feet
Counting the number of incidents in each hexagon....
- Analysis is performed on all hexagons within the bounding polygon feature class.
Evaluating incident counts and number of polygons....
- The aggregation process resulted in 5711 weighted polygons.
- Incident Count Properties:
        Min:        0.0000
        Max:       27.0000
        Mean:       0.5286
        Std. Dev.:  1.6164
***************************** Scale of Analysis ******************************
Looking for an optimal scale of analysis by assessing the intensity of clustering at increasing distances....
- No optimal distance was found using this method.
Determining an optimal distance using the spatial distribution of features....
- The optimal fixed distance band is based on the average distance to 30 nearest neighbors: 2677.0000 US_Feet
***************************** Hot Spot Analysis ******************************
Finding statistically significant clusters of high and low incident counts....
- There are 518 output features statistically significant based on an FDR correction for multiple testing and spatial dependence.
- 0% of features had less than 8 neighbors based on the distance band of 2677.0000 US_Feet
*********************************** Output ***********************************
Creating output feature class: C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb
\Fatal_Injury_inUSD259_OptimizedHotSpotAnalysis
- Red output features represent hot spots where high incident counts cluster.
- Blue output features represent cold spots where low incident counts cluster.
Completed script Optimized Hot Spot Analysis...
```

Figure 75. The Optimized Hot Spot Analysis Report for USD 259

Optimized Outlier Analysis (Spatial Statistics Tools)

Completed Today at 12:08:37 PM

Parameters

Input Features	Fatal_Injury_inUSD259
Output Features	C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb \Fatal_Injury_inUSD259_OptimizedOutlierAnalysis
Analysis Field	
Incident Data Aggregation Method	COUNT_INCIDENTS_WITHIN_HEXAGON_POLYGONS
Bounding Polygons Defining Where Incidents Are Possible	USD_259
Polygons For Aggregating Incidents Into Counts	
Performance Adjustment	BALANCED_499
Cell Size	
Distance Band	

Messages

Start Time: Sunday, May 5, 2019 12:07:44 PM

```
Running script OptimizedOutlierAnalysis...
*************************** Initial Data Assessment ****************************
Making sure there are enough incidents for analysis....
- There are 3019 valid input features.
Looking for locational outliers....
- There were 43 outlier locations; these will not be used to compute the hexagon size.
Evaluating the bounding polygon feature class....
- The total study area is 4245635781.2126 US_Feet.
***************************** Incident Aggregation *****************************
Creating hexagon mesh to use for aggregating incidents....
- Using a hexagon of width 1096.9655 US_Feet and height 950.0000 US_Feet
Counting the number of incidents in each hexagon....
- Analysis is performed on all hexagons within the bounding polygon feature class.
Evaluating incident counts and number of polygons....
- The aggregation process resulted in 5711 weighted polygons.
- Incident Count Properties:
        Min:        0.0000
        Max:       27.0000
        Mean:       0.5286
        Std. Dev.:  1.6164
****************************** Scale of Analysis ******************************
Looking for an optimal scale of analysis by assessing the intensity of clustering at increasing distances....
- No optimal distance was found using this method.
Determining an optimal distance using the spatial distribution of features....
- The optimal fixed distance band is based on the average distance to 30 nearest neighbors: 2677.0000 US_Feet
************************* Optimized Outlier Analysis *************************
Creating the random reference distribution with 499 permutations....
Finding statistically significant outliers of high and low  values....
- There are 1002 output features statistically significant based on an FDR correction for multiple testing and spatial dependence.
- There are 40 statistically significant high outlier features.
- There are 146 statistically significant low outlier features.
- There are 693 features part of statistically significant low clusters.
- There are 123 features part of statistically significant high clusters
- 0% of features had less than 8 neighbors based on the distance band of 2677.0000 US_Feet
************************************ Output ************************************
- Pink output features are part of a cluster of high  values.
- Light Blue output features are part of a cluster of low  values.
- Red output features represent high outliers within a cluster of low  values..
- Blue output features represent low outliers within a cluster of high  values.
Completed script Optimized Outlier Analysis...
```

Figure 76. The Optimized Cluster and Outlier Analysis Report for USD 259

APPENDIX D

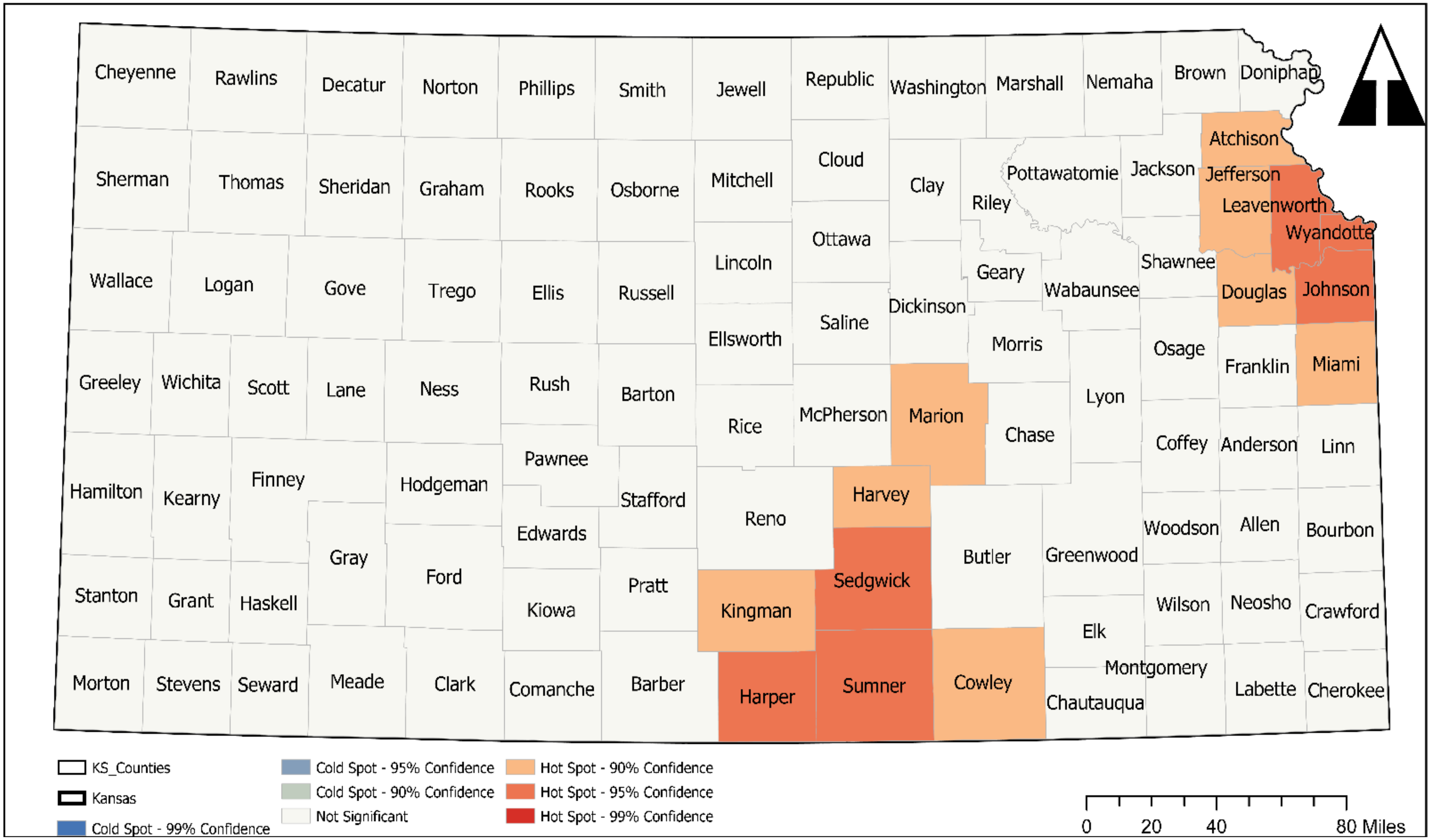
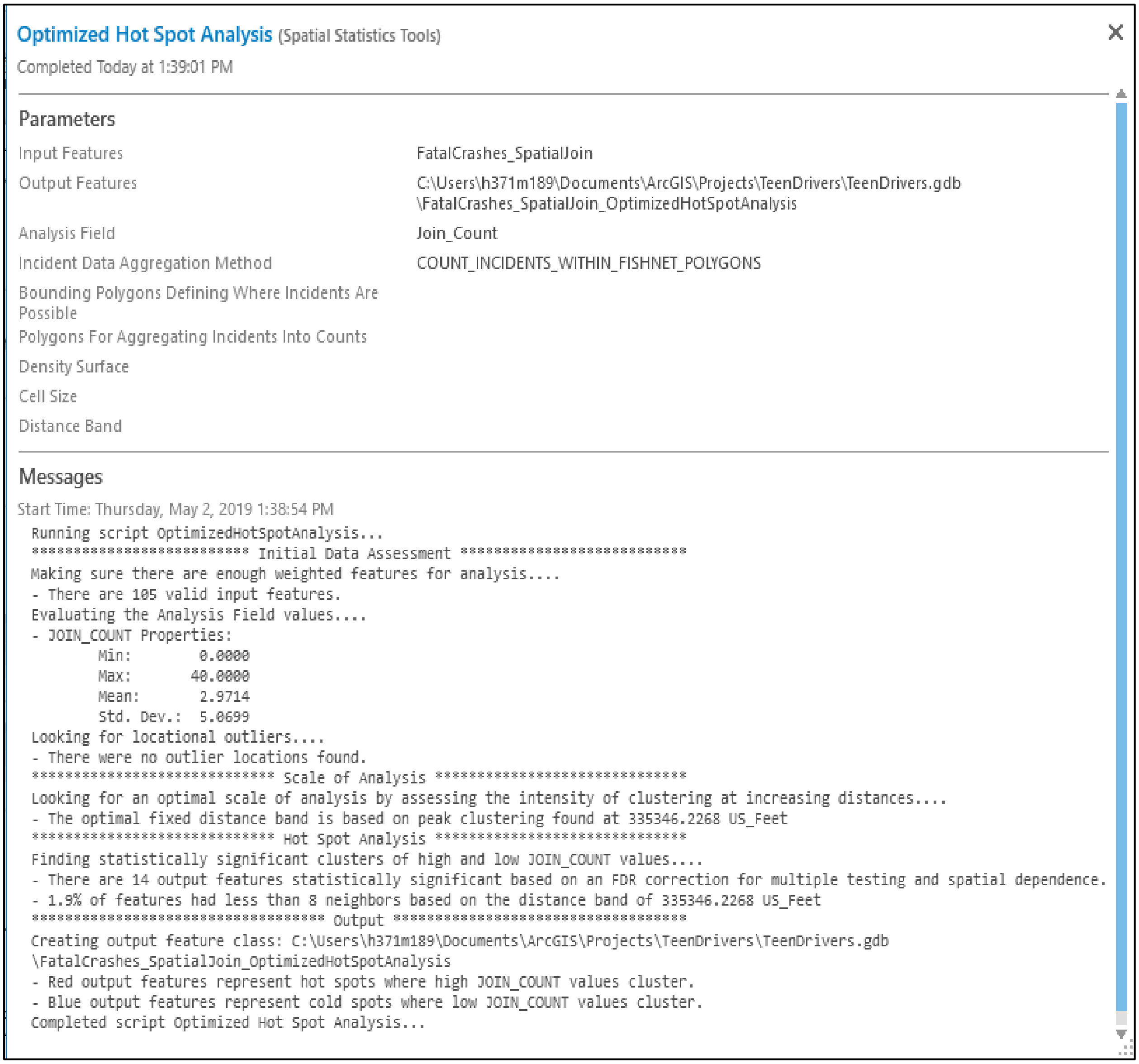


Figure 77. The Optimized Hot Spot Analysis Result for Fatal Crashes on County Levels



Optimized Hot Spot Analysis (Spatial Statistics Tools)

Completed Today at 1:39:01 PM

Parameters

Input Features	FatalCrashes_SpatialJoin
Output Features	C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb \FatalCrashes_SpatialJoin_OptimizedHotSpotAnalysis
Analysis Field	Join_Count
Incident Data Aggregation Method	COUNT_INCIDENTS_WITHIN_FISHNET_POLYGONS
Bounding Polygons Defining Where Incidents Are Possible	
Polygons For Aggregating Incidents Into Counts	
Density Surface	
Cell Size	
Distance Band	

Messages

Start Time: Thursday, May 2, 2019 1:38:54 PM

```
Running script OptimizedHotSpotAnalysis...
************************** Initial Data Assessment ***************************
Making sure there are enough weighted features for analysis....
- There are 105 valid input features.
Evaluating the Analysis Field values....
- JOIN_COUNT Properties:
        Min:        0.0000
        Max:       40.0000
        Mean:       2.9714
        Std. Dev.:  5.0699
Looking for locational outliers....
- There were no outlier locations found.
***************************** Scale of Analysis ******************************
Looking for an optimal scale of analysis by assessing the intensity of clustering at increasing distances....
- The optimal fixed distance band is based on peak clustering found at 335346.2268 US_Feet
***************************** Hot Spot Analysis ******************************
Finding statistically significant clusters of high and low JOIN_COUNT values....
- There are 14 output features statistically significant based on an FDR correction for multiple testing and spatial dependence.
- 1.9% of features had less than 8 neighbors based on the distance band of 335346.2268 US_Feet
*********************************** Output ***********************************
Creating output feature class: C:\Users\h371m189\Documents\ArcGIS\Projects\TeenDrivers\TeenDrivers.gdb
\FatalCrashes_SpatialJoin_OptimizedHotSpotAnalysis
- Red output features represent hot spots where high JOIN_COUNT values cluster.
- Blue output features represent cold spots where low JOIN_COUNT values cluster.
Completed script Optimized Hot Spot Analysis...
```

Figure 78. The Optimized Hot Spot Analysis Report for Fatal Crashes on County Levels

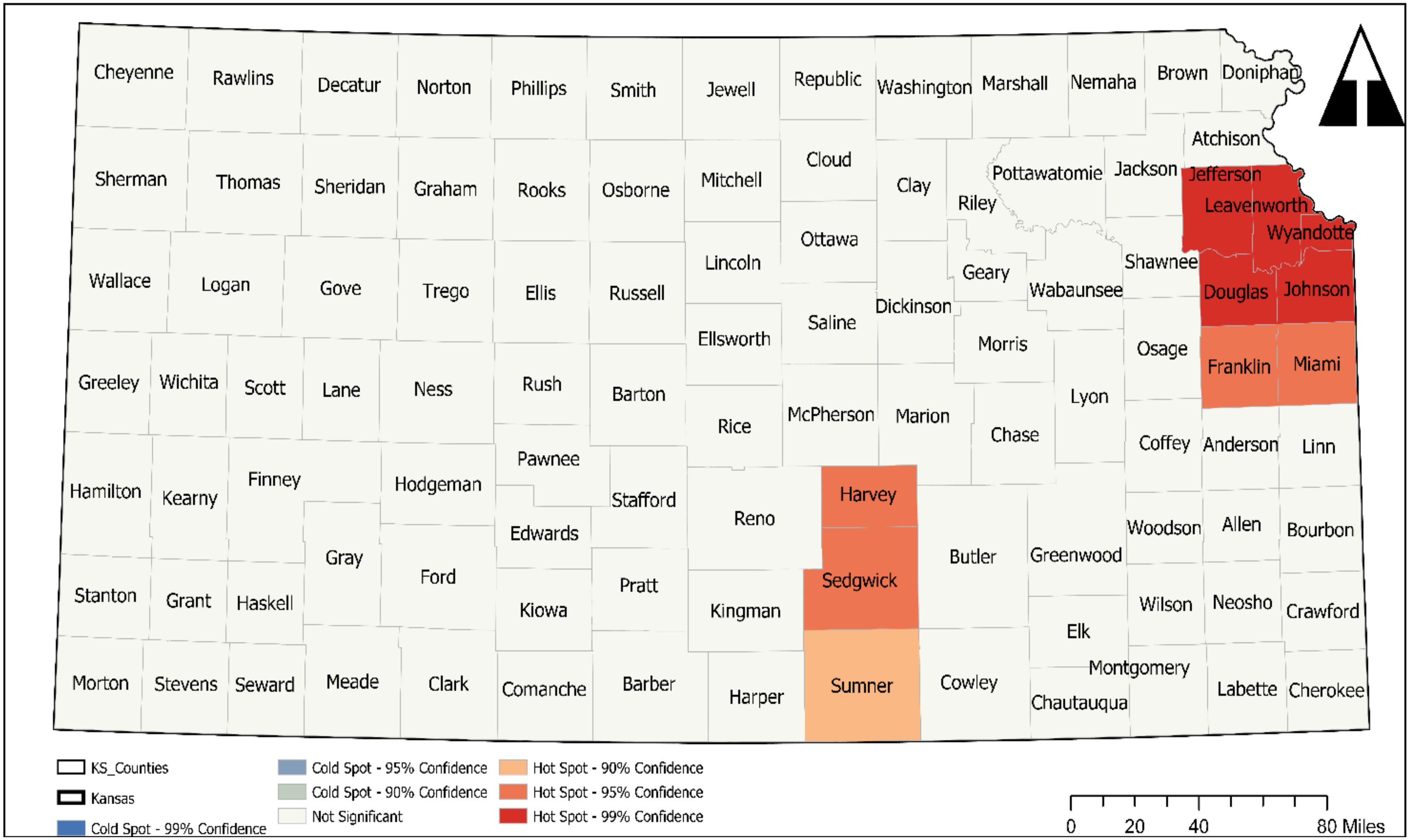


Figure 79. The Optimized Hot Spot Analysis Result for Non-Fatal Crashes on County Levels

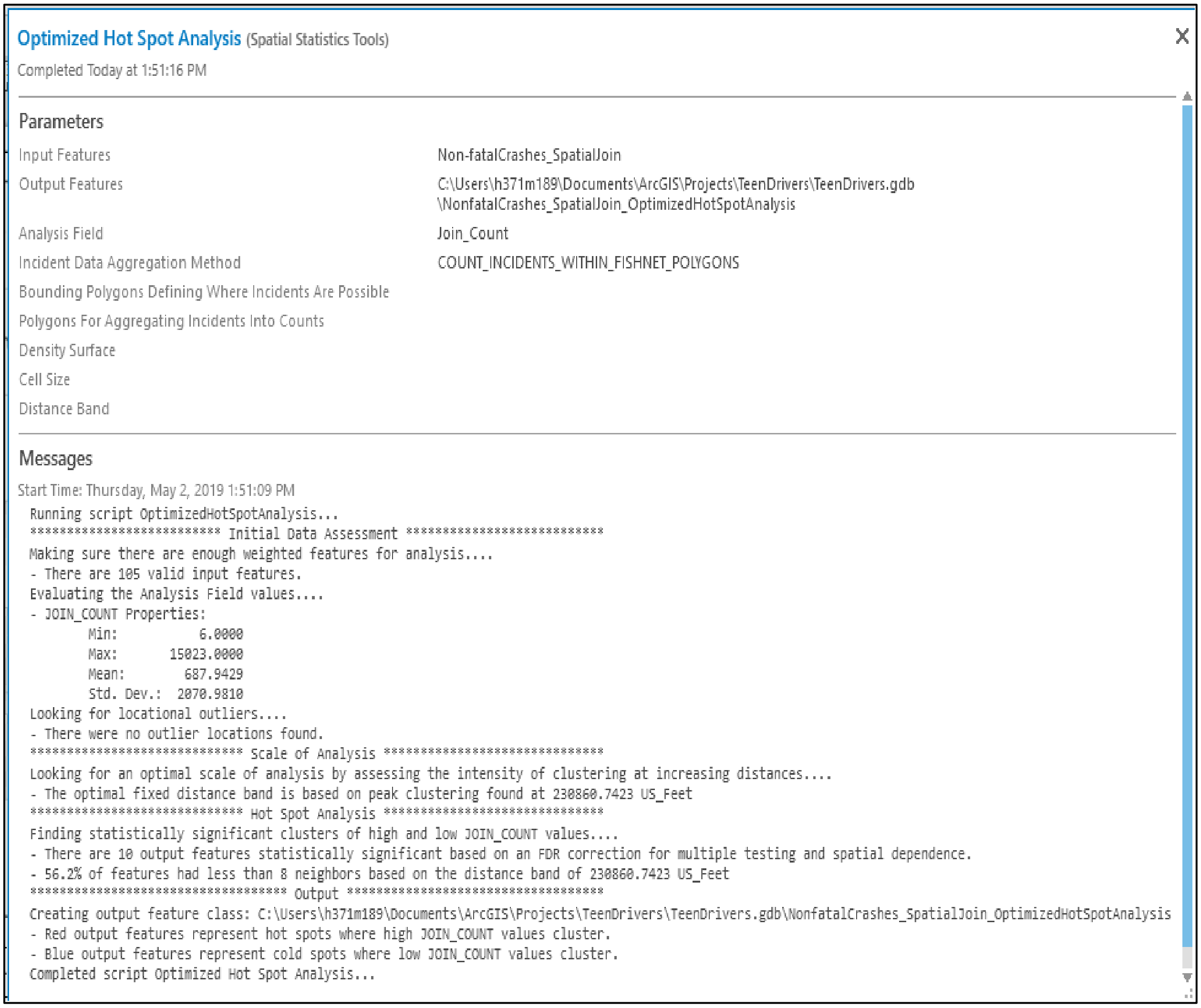


Figure 80. The Optimized Hot Spot Analysis Report for Non-Fatal Crashes on County Levels

```
******************************************************************************
Choose 1 of 18 Summary
     Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)  VIF   SA   Model
 0.90 -398.13 0.00  0.01 1.00 0.65  +LNLOAVG_PC***
 0.87 -369.23 0.00  0.01 1.00 0.00  +LNLOT_POP***
 0.86 -361.07 0.00  0.17 1.00 0.15  +LNPOP***
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
******************************************************************************
Choose 2 of 18 Summary
              Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)   VIF   SA   Model
 0.91 -403.18 0.00  0.01  1.09 0.72  +NONSTATE_RD***  +LNLOAVG_PC***
 0.91 -402.63 0.00  0.01 46.14 0.79  -LNPOP**  +LNLOAVG_PC***
 0.91 -401.31 0.00  0.01 28.69 0.95  -NMALE_OVER15**  +LNLOAVG_PC***
                                   Passing Models
AdjR2    AICc        JB       K(BP)    VIF      SA        Model
0.582000 -243.839212 0.821966 0.040669 1.241102 0.218993  -HIGHSCHOOL***  +AVG_PRECIPT***
******************************************************************************
Choose 3 of 18 Summary
                       Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)   VIF   SA   Model
 0.91 -407.83 0.00  0.00 43.59 0.46  -NMALE_OVER15***  +LNLOT_POP***  +LNLOAVG_PC***
 0.91 -406.83 0.00  0.00 45.33 0.28  -NLABOR_OVER15***  +LNLOT_POP***  +LNLOAVG_PC***
 0.91 -406.69 0.00  0.02 27.00 0.49  -AVG_TRUCK**  +NONSTATE_RD***  +LNLOAVG_PC***
                                             Passing Models
AdjR2    AICc        JB       K(BP)    VIF      SA        Model
0.703382 -278.686388 0.812218 0.015369 1.405761 0.736779  -HIGHSCHOOL***  +AVG_PRECIPT***  +ALL_ROAD***
0.618646 -252.301790 0.381170 0.004374 1.328041 0.884104  -HIGHSCHOOL***  +AVG_PRECIPT***  +NONSTATE_RD**
0.604940 -248.594286 0.276814 0.005884 1.351043 0.449767  -HIGHSCHOOL***  +AV_HOUSEH_INCOM***  +AVG_PRECIPT***
0.560736 -237.457619 0.765416 0.041628 1.160673 0.791590  +AV_HOUSEH_INCOM**  +AVG_PRECIPT***  +ALL_ROAD***
******************************************************************************
Choose 4 of 18 Summary
                               Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)   VIF   SA   Model
 0.92 -411.35 0.00  0.01 45.01 0.76  +NONSTATE_RD***  -NMALE_OVER15***  +LNLOT_POP***  +LNLOAVG_PC***
 0.92 -410.93 0.00  0.01 46.33 0.96  +NONSTATE_RD***  -NLABOR_OVER15***  +LNLOT_POP***  +LNLOAVG_PC***
 0.92 -410.54 0.00  0.01 57.73 0.45  -LNPOP***  +NONSTATE_RD***  +LNLOT_POP***  +LNLOAVG_PC***
                                                       Passing Models
AdjR2    AICc        JB       K(BP)    VIF      SA        Model
0.652948 -260.991956 0.924596 0.000667 1.465990 0.298599  -HIGHSCHOOL***  +AV_HOUSEH_INCOM***  +AVG_PRECIPT***  +NONSTATE_RD***
0.610304 -248.823509 0.610107 0.113525 1.532558 0.200486  +AV_HOUSEH_INCOM***  +UNDER_POV_LEV***  +AVG_PRECIPT***  +ALL_ROAD***
0.562552 -236.686281 0.818961 0.002282 1.528466 0.840563  +AV_HOUSEH_INCOM***  +UNDER_POV_LEV***  +AVG_PRECIPT***  +NONSTATE_RD***
******************************************************************************
Choose 5 of 18 Summary
                                     Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)   VIF   SA   Model
 0.92 -412.30 0.00  0.05 54.90 0.66  +NONSTATE_RD***  +DVMT_ALL*  -NLABOR_OVER15***  +LNLOT_POP***  +LNLOAVG_PC***
 0.92 -412.25 0.00  0.04 51.70 0.97  +NONSTATE_RD***  +DVMT_ALL*  -NMALE_OVER15***  +LNLOT_POP***  +LNLOAVG_PC***
 0.92 -412.10 0.00  0.02 69.56 0.65  -LNPOP***  +NONSTATE_RD***  +DVMT_ALL**  +LNLOT_POP***  +LNLOAVG_PC***
                                                               Passing Models
AdjR2    AICc        JB       K(BP)    VIF      SA        Model
0.716496 -280.985872 0.503282 0.019337 1.717898 0.312782  -HIGHSCHOOL***  +AV_HOUSEH_INCOM**  +UNDER_POV_LEV**  +AVG_PRECIPT***  +ALL_ROAD***
0.669986 -265.035589 0.725813 0.000484 1.763579 0.856484  -HIGHSCHOOL***  +AV_HOUSEH_INCOM***  +UNDER_POV_LEV**  +AVG_PRECIPT***  +NONSTATE_RD***
******************************************************************************
```

Figure 81. The Proposed OLS Models by the Exploratory Regression

APPENDIX E

Table 43. Predicted Number of Crashes by OLS for 2026

County	PC	NONSTATE_RD	Residual	Predicted
Allen	7,225	913.1	0.0226	42
Anderson	5,268	989.9	0.0192	29
Atchison	10,105	756.0	0.0230	59
Barber	2,762	883.7	0.0041	12
Barton	16,760	1,588.7	-0.0013	108
Bourbon	8,445	1,012.2	0.0295	58
Brown	6,395	1,040.7	0.0075	32
Butler	40,529	2,051.9	-0.0164	304
Chase	1,707	530.8	0.0645	12
Chautauqua	1,528	664.0	-0.0004	6
Cherokee	10,866	1,062.0	0.0263	77
Cheyenne	1,932	1,147.9	-0.0299	7
Clark	1,221	618.6	0.0497	7
Clay	5,612	1,096.5	0.0038	27
Cloud	5,724	1,197.9	0.0309	40
Coffey	6,166	1,102.4	0.0116	33
Comanche	1,083	603.1	-0.1232	2
Cowley	20,401	1,419.9	0.0142	165
Crawford	22,259	1,055.9	0.0053	136
Decatur	2,025	1,150.6	-0.1442	3
Dickinson	12,615	1,495.9	-0.0138	62
Doniphan	4,832	582.6	0.0171	22
Douglas	75,823	789.9	0.0062	572
Edwards	1,983	975.8	0.0006	9
Elk	1,541	714.7	0.0251	8
Ellis	17,658	1,425.6	0.0227	157
Ellsworth	3,776	1,014.0	0.0159	20
Finney	23,567	1,279.6	0.0012	151
Ford	19,390	1,511.9	-0.0008	126
Franklin	17,305	999.5	-0.0209	66
Geary	26,952	484.7	-0.0054	104
Gove	1,825	1,111.6	0.0457	13
Graham	1,636	1,182.7	-0.0030	7
Grant	4,512	783.8	0.0335	27
Gray	3,414	1,193.1	0.0116	18
Greeley	871	838.1	-0.0200	3
Greenwood	4,024	1,332.0	-0.0008	19
Hamilton	1,588	825.5	-0.0680	4
Harper	3,268	1,243.7	0.0354	23
Harvey	24,025	972.0	-0.0144	105
Haskell	2,633	819.8	-0.0111	10

Hodgeman	1,152	1,025.1	0.0042	5
Jackson	8,725	1,088.7	0.0160	52
Jefferson	13,529	984.4	-0.0142	55
Jewell	2,147	1,495.7	-0.0492	7
Johnson	468,040	407.4	-0.0313	2126
Kearny	2,536	758.7	0.0443	16
Kingman	5,207	1,317.7	0.0014	26
Kiowa	1,534	804.1	0.0293	8
Labette	12,478	1,102.5	0.0108	74
Lane	1,128	674.9	-0.0679	3
Leavenworth	51,034	914.8	-0.0114	269
Lincoln	2,040	1,038.9	0.0089	10
Linn	7,031	992.2	-0.0227	24
Logan	1,918	836.9	-0.0298	6
Lyon	19,343	1,359.7	0.0077	135
Marion	8,118	1,650.8	-0.0204	36
Marshall	6,886	1,454.7	0.0053	39
McPherson	19,664	1,588.5	-0.0228	94
Meade	2,732	968.5	-0.0069	11
Miami	24,470	1,091.3	-0.0214	102
Mitchell	4,234	1,146.0	0.0254	26
Montgomery	18,859	1,084.5	0.0104	121
Morris	3,699	999.2	0.0223	21
Morton	1,828	669.5	-0.0092	7
Nemaha	6,924	1,265.5	0.0156	42
Neosho	9,295	1,000.0	0.0247	61
Ness	1,940	1,286.0	-0.0022	9
Norton	3,418	1,268.3	0.0474	28
Osage	11,543	1,174.2	-0.0221	44
Osborne	2,459	1,153.1	-0.0231	9
Ottawa	4,150	1,111.1	0.0192	23
Pawnee	3,687	1,281.2	0.0368	27
Phillips	3,667	1,374.0	0.0069	19
Pottawatomie	16,064	1,227.9	-0.0044	85
Pratt	5,621	1,198.4	0.0262	37
Rawlins	1,580	1,198.3	0.0149	8
Reno	40,276	2,236.0	-0.0294	256
Republic	3,280	1,285.8	-0.0248	12
Rice	6,025	1,290.5	-0.0016	29
Riley	34,016	769.9	0.0213	259
Rooks	3,536	1,327.8	0.0179	21
Rush	2,284	1,220.0	0.0127	12
Russell	4,631	1,275.7	0.0301	32
Saline	38,139	1,094.6	-0.0025	239
Scott	3,348	773.8	-0.0042	13
Sedgwick	351,617	1,381.4	-0.0570	1750

Seward	12,665	721.8	0.0195	72
Shawnee	125,296	2,039.4	-0.0186	1339
Sheridan	1,797	1,244.0	0.0031	9
Sherman	3,126	1,203.3	0.0237	19
Smith	2,388	1,434.5	0.0071	12
Stafford	2,616	1,362.6	0.0093	14
Stanton	1,228	761.6	-0.1384	2
Stevens	3,583	1,039.7	-0.0005	16
Sumner	15,035	2,060.2	-0.0233	79
Thomas	4,869	1,488.3	0.0355	39
Trego	2,137	1,144.8	0.0301	13
Wabaunsee	4,830	898.4	0.0276	28
Wallace	933	618.4	-0.0604	2
Washington	3,997	1,535.9	0.0291	29
Wichita	1,353	835.4	0.0415	8
Wilson	5,535	904.7	0.0014	24
Woodson	1,918	740.1	0.0178	9
Wyandotte	101,326	8.1	-0.0395	141
Total				10,824

Table 44. Predicted Number of Crashes by GWR for 2026

County	Passenger Cars	NONSTATE	Intercept	Coeff. NONSTATE	Coeff. PC	Residual	Predicted
Allen	7,225	913.1	-0.85455	0.00954	1.636511	0.015767	43
Anderson	5,268	989.9	-0.85171	0.009195	1.635204	0.010089	29
Atchison	10,105	756.0	-0.83802	0.00755	1.628871	0.013839	58
Barber	2,762	883.7	-0.99637	0.0196	1.730447	0.001036	12
Barton	16,760	1,588.7	-1.00732	0.016405	1.760231	0.000806	108
Bourbon	8,445	1,012.2	-0.85142	0.009519	1.633398	0.024735	58
Brown	6,395	1,040.7	-0.83967	0.007204	1.632171	0.001102	32
Butler	40,529	2,051.9	-0.87945	0.010055	1.659579	0.001827	304
Chase	1,707	530.8	-0.86581	0.009119	1.65009	0.037647	12
Chautauqua	1,528	664.0	-0.88083	0.010772	1.65755	-0.02307	6
Cherokee	10,866	1,062.0	-0.85793	0.010152	1.637121	0.024026	77
Cheyenne	1,932	1,147.9	-1.24305	0.028315	1.95101	-0.02393	7
Clark	1,221	618.6	-1.11682	0.023035	1.840564	0.056533	7
Clay	5,612	1,096.5	-0.86571	0.006782	1.661545	-0.00225	27
Cloud	5,724	1,197.9	-0.89324	0.006329	1.692877	0.027298	40
Coffey	6,166	1,102.4	-0.85528	0.009185	1.638921	0.005891	33
Comanche	1,083	603.1	-1.06676	0.021806	1.793405	-0.12154	2
Cowley	20,401	1,419.9	-0.8945	0.011061	1.670086	0.021359	166
Crawford	22,259	1,055.9	-0.85475	0.009854	1.635251	0.008637	135
Decatur	2,025	1,150.6	-1.20592	0.026592	1.919874	-0.14027	3
Dickinson	12,615	1,495.9	-0.86767	0.007689	1.659144	-0.00756	62
Doniphan	4,832	582.6	-0.83571	0.007332	1.627487	-0.00226	22
Douglas	75,823	789.9	-0.84095	0.008236	1.62871	0.014034	573
Edwards	1,983	975.8	-1.07669	0.021912	1.804068	0.00097	9
Elk	1,541	714.7	-0.87174	0.010244	1.650793	0.00216	8
Ellis	17,658	1,425.6	-1.10008	0.021094	1.834446	0.019225	158
Ellsworth	3,776	1,014.0	-0.94046	0.009584	1.725849	0.008368	20
Finney	23,567	1,279.6	-1.18258	0.025658	1.898387	-0.00699	151
Ford	19,390	1,511.9	-1.12734	0.023637	1.849054	-0.00549	126
Franklin	17,305	999.5	-0.84248	0.008569	1.628737	-0.02014	66
Geary	26,952	484.7	-0.85719	0.007682	1.648194	-0.01212	107
Gove	1,825	1,111.6	-1.18589	0.025947	1.900983	0.05035	13
Graham	1,636	1,182.7	-1.16503	0.02461	1.885985	0.000458	7
Grant	4,512	783.8	-1.19694	0.025716	1.913399	0.038028	27
Gray	3,414	1,193.1	-1.16302	0.024854	1.881215	0.012956	18
Greeley	871	838.1	-1.2263	0.027276	1.937687	-0.00348	3
Greenwood	4,024	1,332.0	-0.86581	0.00972	1.6472	-0.00649	19

Hamilton	1,588	825.5	-1.21969	0.026789	1.932817	-0.0564	4
Harper	3,268	1,243.7	-0.94727	0.015212	1.702347	0.031103	23
Harvey	24,025	972.0	-0.89326	0.009914	1.674469	-0.01261	105
Haskell	2,633	819.8	-1.18025	0.025202	1.897945	-0.00464	10
Hodgeman	1,152	1,025.1	-1.1348	0.024034	1.855332	0.009147	5
Jackson	8,725	1,088.7	-0.84187	0.007503	1.633095	0.013268	52
Jefferson	13,529	984.4	-0.84104	0.007897	1.630355	-0.01449	55
Jewell	2,147	1,495.7	-0.97383	0.008512	1.766959	-0.04988	7
Johnson	468,040	407.4	-0.83761	0.008279	1.62507	-0.01904	2,213
Kearny	2,536	758.7	-1.20461	0.026318	1.918839	0.052614	16
Kingman	5,207	1,317.7	-0.94127	0.014903	1.697774	-0.00031	26
Kiowa	1,534	804.1	-1.07162	0.021611	1.799918	0.030316	8
Labette	12,478	1,102.5	-0.86214	0.010244	1.640994	0.010443	74
Lane	1,128	674.9	-1.17774	0.025681	1.893308	-0.05582	3
Leavenworth	51,034	914.8	-0.83967	0.008035	1.628289	-0.00696	255
Lincoln	2,040	1,038.9	-0.94425	0.007866	1.739155	-0.00083	10
Linn	7,031	992.2	-0.8482	0.009163	1.631759	-0.02954	24
Logan	1,918	836.9	-1.21172	0.026944	1.92372	-0.02047	6
Lyon	19,343	1,359.7	-0.85545	0.008803	1.640939	0.014628	135
Marion	8,118	1,650.8	-0.8756	0.008992	1.660859	-0.01592	36
Marshall	6,886	1,454.7	-0.85002	0.006709	1.645327	0.007308	39
McPherson	19,664	1,588.5	-0.894	0.008859	1.680725	-0.01375	95
Meade	2,732	968.5	-1.15063	0.023894	1.872492	-0.00295	11
Miami	24,470	1,091.3	-0.83927	0.008585	1.625369	-0.01658	102
Mitchell	4,234	1,146.0	-0.95501	0.007478	1.75262	0.020357	26
Montgomery	18,859	1,084.5	-0.8683	0.010392	1.646583	0.013442	122
Morris	3,699	999.2	-0.85703	0.008203	1.645529	0.01108	21
Morton	1,828	669.5	-1.20554	0.025606	1.923155	0.002951	7
Nemaha	6,924	1,265.5	-0.84476	0.007096	1.637954	0.014179	42
Neosho	9,295	1,000.0	-0.85838	0.009898	1.638754	0.020384	61
Ness	1,940	1,286.0	-1.14398	0.024299	1.864154	-0.00066	9
Norton	3,418	1,268.3	-1.17657	0.024606	1.898689	0.047033	28
Osage	11,543	1,174.2	-0.8437	0.008353	1.631037	-0.0215	44
Osborne	2,459	1,153.1	-1.04095	0.014291	1.807916	-0.02584	9
Ottawa	4,150	1,111.1	-0.88491	0.006276	1.684619	0.011672	23
Pawnee	3,687	1,281.2	-1.07294	0.021202	1.804102	0.035826	27
Phillips	3,667	1,374.0	-1.13182	0.020848	1.870522	0.005545	19
Pottawatomie	16,064	1,227.9	-0.84717	0.007266	1.639759	-0.00032	85
Pratt	5,621	1,198.4	-1.00243	0.018905	1.740906	0.025364	37
Rawlins	1,580	1,198.3	-1.22706	0.027682	1.936988	0.020982	8
Reno	40,276	2,236.0	-0.9271	0.013025	1.693002	-0.01423	256
Republic	3,280	1,285.8	-0.89946	0.006125	1.700351	-0.02957	12
Rice	6,025	1,290.5	-0.93766	0.011545	1.712277	-0.00345	29
Riley	34,016	769.9	-0.85345	0.007124	1.647022	0.022617	264

Rooks	3,536	1,327.8	-1.11656	0.020966	1.85333	0.016634	21
Rush	2,284	1,220.0	-1.08842	0.021451	1.819703	0.012216	12
Russell	4,631	1,275.7	-1.02187	0.014961	1.783752	0.028246	32
Saline	38,139	1,094.6	-0.88677	0.007316	1.681219	0.002368	239
Scott	3,348	773.8	-1.19759	0.026386	1.911125	0.002029	13
Sedgwick	351,617	1,381.4	-0.90373	0.010887	1.680274	-0.03367	1,890
Seward	12,665	721.8	-1.17299	0.024561	1.893183	0.019894	72
Shawnee	125,296	2,039.4	-0.84544	0.008048	1.6342	-0.00392	983
Sheridan	1,797	1,244.0	-1.19519	0.026312	1.909465	0.006781	9
Sherman	3,126	1,203.3	-1.23531	0.027916	1.944487	0.024786	19
Smith	2,388	1,434.5	-1.06422	0.014469	1.832083	0.006921	12
Stafford	2,616	1,362.6	-0.99699	0.01776	1.741612	0.005694	14
Stanton	1,228	761.6	-1.21177	0.026166	1.927215	-0.12529	2
Stevens	3,583	1,039.7	-1.19045	0.025122	1.909233	0.002645	16
Sumner	15,035	2,060.2	-0.91769	0.011928	1.689321	-0.01335	79
Thomas	4,869	1,488.3	-1.21744	0.027276	1.928439	0.030826	39
Trego	2,137	1,144.8	-1.15042	0.024229	1.871798	0.032273	13
Wabaunsee	4,830	898.4	-0.85048	0.008036	1.639476	0.016331	28
Wallace	933	618.4	-1.22867	0.02749	1.939283	-0.04114	2
Washington	3,997	1,535.9	-0.86312	0.006202	1.661598	0.028768	29
Wichita	1,353	835.4	-1.21323	0.026889	1.925477	0.053749	8
Wilson	5,535	904.7	-0.86317	0.009996	1.643194	-0.00765	24
Woodson	1,918	740.1	-0.85874	0.009598	1.640531	-0.00399	9
Wyandotte	101,326	8.1	-0.83844	0.008155	1.626458	-0.02862	240
Total							10,795

Table 45. Growth Rates and Predicted Values for the Exploratory Variables

No.	County	Non-state Rd. (miles) Growth Rate (%)	Non-state Rd. (miles) Predicted in 2026	Passenger Cars (No.) Growth Rate (%)	Passenger Cars (No.) PC in 2026
1	Allen	0.091	913.07	-0.817	7225.4
2	Anderson	0.057	989.89	-0.002	5268.0
3	Atchison	0.400	755.96	0.006	10104.9
4	Barber	-0.012	883.69	-0.067	2762.3
5	Barton	-0.119	1588.73	0.022	16760.0
6	Bourbon	0.027	1012.17	0.005	8445.5
7	Brown	0.006	1040.68	0.010	6395.2
8	Butler	-0.065	2051.89	-0.006	40528.8
9	Chase	0.004	530.77	-0.001	1706.9
10	Chautauqua	0.178	664.02	0.128	1528.5
11	Cherokee	0.214	1061.98	0.011	10865.7
12	Cheyenne	-0.007	1147.91	0.012	1932.2
13	Clark	-0.017	618.55	0.067	1221.2
14	Clay	0.007	1096.48	0.020	5612.0
15	Cloud	0.015	1197.89	0.027	5724.2
16	Coffey	0.041	1102.43	-0.001	6165.6
17	Comanche	-0.054	603.13	0.073	1082.9
18	Cowley	-0.294	1419.89	0.007	20400.8
19	Crawford	0.050	1055.93	-0.001	22258.9
20	Decatur	-0.052	1150.55	-0.010	2025.0
21	Dickinson	0.004	1495.93	0.015	12615.4
22	Doniphan	0.174	582.58	0.002	4831.8
23	Douglas	-0.001	789.92	-0.131	75822.8
24	Edwards	-0.003	975.76	0.000	1982.9
25	Elk	0.074	714.69	0.061	1541.4
26	Ellis	0.645	1425.55	0.027	17657.6
27	Ellsworth	-0.005	1013.99	0.002	3775.7
28	Finney	-0.082	1279.63	-0.065	23567.2
29	Ford	0.038	1511.94	-0.001	19389.8
30	Franklin	0.013	999.54	-0.010	17304.9
31	Geary	-0.757	484.75	-0.001	26952.3
32	Gove	0.000	1111.58	0.000	1825.0
33	Graham	0.096	1182.69	0.070	1636.4
34	Grant	0.067	783.83	0.011	4512.2
35	Gray	0.011	1193.14	0.012	3414.2
36	Greeley	-0.017	838.07	0.101	870.8
37	Greenwood	0.003	1332.03	0.010	4024.2
38	Hamilton	0.004	825.49	-0.008	1587.8
39	Harper	-0.098	1243.73	-0.075	3268.3

40	Harvey	-0.109	972.00	-0.004	24025.2
41	Haskell	0.012	819.78	-0.003	2633.3
42	Hodgeman	0.034	1025.13	0.120	1151.7
43	Jackson	-0.065	1088.67	-0.005	8724.5
44	Jefferson	-0.214	984.42	0.007	13529.1
45	Jewell	-0.009	1495.68	-0.028	2147.0
46	Johnson	-0.885	407.38	-0.029	468039.8
47	Kearny	0.005	758.69	0.000	2536.0
48	Kingman	0.009	1317.75	0.077	5207.0
49	Kiowa	-0.042	804.10	-0.009	1533.6
50	Labette	0.113	1102.54	-0.053	12478.2
51	Lane	0.004	674.87	0.045	1128.1
52	Leavenworth	1.652	914.79	-0.028	51034.3
53	Lincoln	0.001	1038.91	-0.001	2039.7
54	Linn	0.104	992.18	-0.080	7031.4
55	Logan	-0.033	836.95	-0.003	1918.3
56	Lyon	0.031	1359.75	-0.024	19342.9
57	Marion	0.017	1650.76	-0.006	8118.4
58	Marshall	-0.044	1454.74	0.018	6886.4
59	McPherson	-0.006	1588.47	0.003	19664.0
60	Meade	0.088	968.49	0.025	2731.9
61	Miami	0.124	1091.27	-0.019	24470.1
62	Mitchell	0.006	1145.99	-0.004	4234.5
63	Montgomery	-0.241	1084.48	0.016	18858.8
64	Morris	0.000	999.23	0.002	3698.7
65	Morton	0.149	669.52	0.043	1827.8
66	Nemaha	-0.001	1265.54	0.004	6923.6
67	Neosho	0.264	1000.03	0.029	9294.6
68	Ness	0.011	1286.03	0.030	1939.8
69	Norton	0.097	1268.28	0.007	3418.4
70	Osage	-0.105	1174.15	0.011	11543.4
71	Osborne	0.000	1153.05	0.010	2458.6
72	Ottawa	-0.011	1111.09	0.016	4149.7
73	Pawnee	0.038	1281.18	0.038	3686.9
74	Phillips	0.032	1374.02	0.014	3667.1
75	Pottawatomi	0.113	1227.92	-0.031	16063.8
76	Pratt	0.012	1198.41	-0.005	5620.9
77	Rawlins	0.045	1198.29	0.060	1580.4
78	Reno	-0.055	2236.03	0.006	40275.9
79	Republic	-0.001	1285.84	0.003	3279.9
80	Rice	0.263	1290.45	0.028	6025.0
81	Riley	-0.579	769.94	-0.004	34016.1
82	Rooks	-0.001	1327.82	-0.002	3536.4
83	Rush	0.012	1219.95	0.021	2283.9
84	Russell	-0.130	1275.70	0.006	4630.7
85	Saline	0.027	1094.61	-0.009	38138.7

86	Scott	-0.033	773.76	0.001	3348.5
87	Sedgwick	-2.181	1381.40	0.012	351617.3
88	Seward	0.002	721.80	0.000	12664.5
89	Shawnee	7.235	2039.45	-0.006	125295.6
90	Sheridan	-0.001	1244.02	-0.001	1796.8
91	Sherman	0.005	1203.30	0.085	3126.5
92	Smith	0.131	1434.48	0.017	2388.1
93	Stafford	0.030	1362.57	0.023	2616.1
94	Stanton	-0.260	761.58	0.094	1227.5
95	Stevens	-0.013	1039.70	0.025	3583.0
96	Sumner	0.057	2060.24	0.003	15035.0
97	Thomas	0.063	1488.31	0.002	4869.1
98	Trego	-0.025	1144.80	0.029	2137.3
99	Wabaunsee	0.015	898.35	0.015	4830.1
100	Wallace	-0.062	618.42	0.065	933.1
101	Washington	-0.043	1535.90	0.008	3997.1
102	Wichita	0.079	835.45	0.131	1352.6
103	Wilson	-0.213	904.69	0.068	5535.5
104	Woodson	-0.071	740.13	0.006	1918.2
105	Wyandotte	-1.176	7.13	-0.118	101325.8

Table 46. Validation of the OLS Model for 2017

County	PC (No.)	Non-State (miles)	Residual	Predicted	Occurred	Variation
Allen	7,700	907.60	0.022641	46	37	23.1%
Anderson	5,206	988.47	0.019184	29	40	-27.9%
Atchison	10,100	731.19	0.022953	58	58	-0.4%
Barber	2,643	876.27	0.004076	11	10	13.9%
Barton	16,210	1612.88	-0.00127	104	93	12.0%
Bourbon	8,258	1008.39	0.029455	57	61	-7.1%
Brown	6,411	1041.06	0.007451	32	28	14.7%
Butler	40,962	2065.76	-0.016411	310	298	3.9%
Chase	1,712	532.11	0.064543	12	18	-35.0%
Chautauqua	1,446	657.38	-0.000378	6	12	-52.6%
Cherokee	10,468	1039.70	0.026331	73	78	-6.2%
Cheyenne	1,871	1148.68	-0.029899	6	6	7.3%
Clark	1,159	616.18	0.049676	7	7	0.4%
Clay	5,426	1098.19	0.003848	26	28	-6.8%
Cloud	5,600	1197.62	0.030882	39	51	-23.0%
Coffey	6,143	1105.43	0.011606	33	30	10.1%
Comanche	1,036	604.45	-0.123245	2	3	-39.7%
Cowley	20,244	1474.11	0.014248	167	148	12.7%
Crawford	22,281	1050.88	0.005287	135	166	-18.4%
Decatur	2,005	1175.66	-0.144163	3	6	-53.1%
Dickinson	12,669	1500.14	-0.013794	62	53	17.3%
Doniphan	4,832	576.27	0.017139	22	12	80.7%
Douglas	77,186	806.43	0.006227	592	579	2.3%
Edwards	1,978	975.25	0.000571	9	9	-5.2%
Elk	1,538	714.37	0.025097	8	1	670.8%
Ellis	17,427	1347.56	0.02265	149	129	15.9%
Ellsworth	3,798	1017.95	0.015855	20	22	-10.0%
Finney	23,704	1304.36	0.001197	154	137	12.2%
Ford	19,192	1512.86	-0.000763	125	161	-22.6%
Franklin	17,572	1004.16	-0.020917	68	65	3.9%
Geary	26,047	548.52	-0.005424	104	91	14.8%
Gove	1,839	1120.28	0.045657	13	17	-24.0%
Graham	1,556	1180.86	-0.002992	7	5	38.7%
Grant	4,491	779.32	0.03346	27	31	-14.0%
Gray	3,342	1191.81	0.011601	17	16	8.9%
Greeley	842	839.48	-0.020005	3	3	7.0%
Greenwood	3,946	1348.05	-0.00082	19	16	18.7%
Hamilton	1,554	830.09	-0.067989	4	8	-53.6%
Harper	3,214	1246.47	0.035438	22	19	17.2%

Harvey	23,802	988.22	-0.014434	105	115	-8.8%
Haskell	2,599	819.77	-0.011145	9	9	5.1%
Hodgeman	1,094	1021.32	0.004208	5	8	-36.4%
Jackson	8,614	1105.80	0.016034	52	55	-6.1%
Jefferson	13,671	1044.87	-0.014202	57	45	26.1%
Jewell	2,080	1493.94	-0.049186	6	7	-7.3%
Johnson	472,138	477.86	-0.031293	2327	2224	4.6%
Kearny	2,550	769.24	0.044346	16	13	23.8%
Kingman	5,127	1316.62	0.001432	26	24	6.8%
Kiowa	1,515	807.46	0.029343	8	7	16.4%
Labette	12,358	1092.35	0.010848	73	69	6.1%
Lane	1,039	676.80	-0.067863	3	4	-34.7%
Leavenworth	50,982	779.04	-0.011419	247	268	-7.7%
Lincoln	1,980	1045.95	0.008898	9	12	-21.3%
Linn	7,210	992.05	-0.022732	25	25	-0.8%
Logan	1,882	842.50	-0.029774	6	6	-0.7%
Lyon	19,427	1359.88	0.0077	136	155	-12.3%
Marion	8,109	1651.73	-0.020445	36	46	-21.8%
Marshall	6,774	1465.01	0.00533	39	39	-1.1%
McPherson	19,697	1598.92	-0.02278	95	91	4.2%
Meade	2,574	973.33	-0.006913	10	8	28.3%
Miami	24,340	1080.25	-0.021401	101	130	-22.2%
Mitchell	4,215	1149.17	0.025392	26	30	-13.4%
Montgomery	18,239	1126.83	0.010367	119	128	-7.4%
Morris	3,671	1015.13	0.022335	21	19	8.0%
Morton	1,674	646.76	-0.00921	6	8	-24.9%
Nemaha	6,951	1268.77	0.015576	42	46	-7.7%
Neosho	9,147	977.63	0.024673	59	48	23.2%
Ness	1,873	1284.57	-0.002163	9	8	7.3%
Norton	3,346	1261.87	0.047427	27	24	12.5%
Osage	11,282	1189.75	-0.022101	44	26	67.6%
Osborne	2,464	1159.93	-0.023099	9	5	77.7%
Ottawa	4,041	1116.96	0.019178	23	25	-8.9%
Pawnee	3,529	1275.74	0.036821	25	19	33.7%
Phillips	3,640	1376.65	0.006891	19	17	12.7%
Pottawatomie	16,305	1230.16	-0.00437	87	86	0.6%
Pratt	5,519	1200.11	0.026182	36	40	-9.1%
Rawlins	1,462	1198.24	0.014927	8	5	54.3%
Reno	39,572	2248.70	-0.029412	251	252	-0.2%
Republic	3,232	1287.51	-0.024832	12	15	-21.5%
Rice	5,911	1262.32	-0.001606	28	26	9.2%
Riley	33,784	828.67	0.021307	266	273	-2.7%

Rooks	3,458	1335.58	0.0179	20	13	56.0%
Rush	2,302	1219.82	0.012709	12	20	-39.9%
Russell	4,610	1303.30	0.030132	32	21	53.1%
Saline	37,729	1096.13	-0.002541	236	262	-10.1%
Scott	3,317	776.41	-0.004207	13	15	-15.4%
Sedgwick	350,456	1721.20	-0.057026	2086	2273	-8.2%
Seward	12,163	721.67	0.019527	68	74	-7.8%
Shawnee	126,545	1012.08	-0.018587	805	785	2.5%
Sheridan	1,734	1246.75	0.003069	8	11	-24.7%
Sherman	3,187	1202.69	0.02374	19	10	90.2%
Smith	2,318	1418.25	0.007112	12	15	-19.6%
Stafford	2,626	1360.37	0.009256	14	12	15.2%
Stanton	1,184	785.07	-0.138429	2	0	N/A
Stevens	3,342	1033.85	-0.000472	15	3	384.6%
Sumner	14,997	2048.75	-0.023256	79	80	-1.4%
Thomas	4,862	1488.99	0.035512	39	30	29.7%
Trego	2,096	1158.68	0.030076	13	14	-8.6%
Wabaunsee	4,729	904.41	0.027598	28	23	20.6%
Wallace	924	620.60	-0.060417	2	2	23.7%
Washington	3,909	1543.23	0.029144	28	22	28.6%
Wichita	1,233	828.91	0.041538	7	6	24.6%
Wilson	5,202	930.92	0.00143	23	27	-15.6%
Woodson	1,908	749.09	0.017831	9	10	-9.8%
Wyandotte	104,129	5.89	-0.03948	145	472	-69.3%
Total	1,971,987	114797.26		10,801	11,212	

Table 47. Validation of the GWR Model for 2017

County	PC	Non-state	Intercept	Coeff. (Non-state)	Coeff. (PC)	Residual	Predicted	Occurred	Variation
Allen	7,700	907.60	-0.8546	0.0095	1.6365	0.0158	46	37	23.47%
Anderson	5,206	988.47	-0.8517	0.0092	1.6352	0.0101	29	40	-27.92%
Atchison	10,100	731.19	-0.8380	0.0076	1.6289	0.0138	58	58	-0.54%
Barber	2,643	876.27	-0.9964	0.0196	1.7304	0.0010	11	10	14.27%
Barton	16,210	1612.88	-1.0073	0.0164	1.7602	0.0008	104	93	12.19%
Bourbon	8,258	1008.39	-0.8514	0.0095	1.6334	0.0247	57	61	-6.79%
Brown	6,411	1041.06	-0.8397	0.0072	1.6322	0.0011	32	28	14.79%
Butler	40,962	2065.76	-0.8795	0.0101	1.6596	0.0018	308	298	3.48%
Chase	1,712	532.11	-0.8658	0.0091	1.6501	0.0376	12	18	-34.97%
Chautauqua	1,446	657.38	-0.8808	0.0108	1.6576	-0.0231	6	12	-52.20%
Cherokee	10,468	1039.70	-0.8579	0.0102	1.6371	0.0240	73	78	-5.81%
Cheyenne	1,871	1148.68	-1.2430	0.0283	1.9510	-0.0239	6	6	6.94%
Clark	1,159	616.18	-1.1168	0.0230	1.8406	0.0565	7	7	0.29%
Clay	5,426	1098.19	-0.8657	0.0068	1.6615	-0.0023	26	28	-6.67%
Cloud	5,600	1197.62	-0.8932	0.0063	1.6929	0.0273	39	51	-22.90%
Coffey	6,143	1105.43	-0.8553	0.0092	1.6389	0.0059	33	30	10.11%
Comanche	1,036	604.45	-1.0668	0.0218	1.7934	-0.1215	2	3	-39.68%
Cowley	20,244	1474.11	-0.8945	0.0111	1.6701	0.0214	167	148	12.77%
Crawford	22,281	1050.88	-0.8548	0.0099	1.6353	0.0086	135	166	-18.55%
Decatur	2,005	1175.66	-1.2059	0.0266	1.9199	-0.1403	3	6	-53.13%
Dickinson	12,669	1500.14	-0.8677	0.0077	1.6591	-0.0076	62	53	17.36%
Doniphan	4,832	576.27	-0.8357	0.0073	1.6275	-0.0023	22	12	80.72%
Douglas	77,186	806.43	-0.8410	0.0082	1.6287	0.0140	588	579	1.63%
Edwards	1,978	975.25	-1.0767	0.0219	1.8041	0.0010	9	9	-5.19%
Elk	1,538	714.37	-0.8717	0.0102	1.6508	0.0022	8	1	672.37%
Ellis	17,427	1347.56	-1.1001	0.0211	1.8344	0.0192	149	129	15.85%
Ellsworth	3,798	1017.95	-0.9405	0.0096	1.7258	0.0084	20	22	-9.97%

Finney	23,704	1304.36	-1.1826	0.0257	1.8984	-0.0070	154	137	12.39%
Ford	19,192	1512.86	-1.1273	0.0236	1.8491	-0.0055	125	161	-22.61%
Franklin	17,572	1004.16	-0.8425	0.0086	1.6287	-0.0201	67	65	3.55%
Geary	26,047	548.52	-0.8572	0.0077	1.6482	-0.0121	105	91	15.00%
Gove	1,839	1120.28	-1.1859	0.0259	1.9010	0.0504	13	17	-24.13%
Graham	1,556	1180.86	-1.1650	0.0246	1.8860	0.0005	7	5	38.27%
Grant	4,491	779.32	-1.1969	0.0257	1.9134	0.0380	27	31	-14.03%
Gray	3,342	1191.81	-1.1630	0.0249	1.8812	0.0130	17	16	8.88%
Greeley	842	839.48	-1.2263	0.0273	1.9377	-0.0035	3	3	6.76%
Greenwood	3,946	1348.05	-0.8658	0.0097	1.6472	-0.0065	19	16	18.89%
Hamilton	1,554	830.09	-1.2197	0.0268	1.9328	-0.0564	4	8	-53.57%
Harper	3,214	1246.47	-0.9473	0.0152	1.7023	0.0311	22	19	17.56%
Harvey	23,802	988.22	-0.8933	0.0099	1.6745	-0.0126	105	115	-8.87%
Haskell	2,599	819.77	-1.1803	0.0252	1.8979	-0.0046	9	9	5.02%
Hodgeman	1,094	1021.32	-1.1348	0.0240	1.8553	0.0091	5	8	-36.48%
Jackson	8,614	1105.80	-0.8419	0.0075	1.6331	0.0133	52	55	-6.06%
Jefferson	13,671	1044.87	-0.8410	0.0079	1.6304	-0.0145	57	45	25.96%
Jewell	2,080	1493.94	-0.9738	0.0085	1.7670	-0.0499	6	7	-7.22%
Johnson	472,138	477.86	-0.8376	0.0083	1.6251	-0.0190	2,314	2224	4.06%
Kearny	2,550	769.24	-1.2046	0.0263	1.9188	0.0526	16	13	23.75%
Kingman	5,127	1316.62	-0.9413	0.0149	1.6978	-0.0003	26	24	7.01%
Kiowa	1,515	807.46	-1.0716	0.0216	1.7999	0.0303	8	7	16.46%
Labette	12,358	1092.35	-0.8621	0.0102	1.6410	0.0104	73	69	6.30%
Lane	1,039	676.80	-1.1777	0.0257	1.8933	-0.0558	3	4	-35.00%
Leavenworth	50,982	779.04	-0.8397	0.0080	1.6283	-0.0070	246	268	-8.05%
Lincoln	1,980	1045.95	-0.9443	0.0079	1.7392	-0.0008	9	12	-21.21%
Linn	7,210	992.05	-0.8482	0.0092	1.6318	-0.0295	25	25	-1.33%
Logan	1,882	842.50	-1.2117	0.0269	1.9237	-0.0205	6	6	-0.74%
Lyon	19,427	1359.88	-0.8554	0.0088	1.6409	0.0146	136	155	-12.44%
Marion	8,109	1651.73	-0.8756	0.0090	1.6609	-0.0159	36	46	-21.70%

Marshall	6,774	1465.01	-0.8500	0.0067	1.6453	0.0073	39	39	-0.94%
McPherson	19,697	1598.92	-0.8940	0.0089	1.6807	-0.0137	95	91	4.22%
Meade	2,574	973.33	-1.1506	0.0239	1.8725	-0.0029	10	8	28.10%
Miami	24,340	1080.25	-0.8393	0.0086	1.6254	-0.0166	101	130	-22.36%
Mitchell	4,215	1149.17	-0.9550	0.0075	1.7526	0.0204	26	30	-13.32%
Montgomery	18,239	1126.83	-0.8683	0.0104	1.6466	0.0134	119	128	-6.96%
Morris	3,671	1015.13	-0.8570	0.0082	1.6455	0.0111	21	19	8.13%
Morton	1,674	646.76	-1.2055	0.0256	1.9232	0.0030	6	8	-25.20%
Nemaha	6,951	1268.77	-0.8448	0.0071	1.6380	0.0142	43	46	-7.56%
Neosho	9,147	977.63	-0.8584	0.0099	1.6388	0.0204	59	48	23.59%
Ness	1,873	1284.57	-1.1440	0.0243	1.8642	-0.0007	9	8	7.03%
Norton	3,346	1261.87	-1.1766	0.0246	1.8987	0.0470	27	24	12.38%
Osage	11,282	1189.75	-0.8437	0.0084	1.6310	-0.0215	44	26	67.70%
Osborne	2,464	1159.93	-1.0409	0.0143	1.8079	-0.0258	9	5	77.71%
Ottawa	4,041	1116.96	-0.8849	0.0063	1.6846	0.0117	23	25	-8.71%
Pawnee	3,529	1275.74	-1.0729	0.0212	1.8041	0.0358	25	19	33.72%
Phillips	3,640	1376.65	-1.1318	0.0208	1.8705	0.0055	19	17	12.64%
Pottawatomi.	16,305	1230.16	-0.8472	0.0073	1.6398	-0.0003	86	86	0.21%
Pratt	5,519	1200.11	-1.0024	0.0189	1.7409	0.0254	36	40	-8.96%
Rawlins	1,462	1198.24	-1.2271	0.0277	1.9370	0.0210	8	5	52.95%
Reno	39,572	2248.70	-0.9271	0.0130	1.6930	-0.0142	252	252	-0.13%
Republic	3,232	1287.51	-0.8995	0.0061	1.7004	-0.0296	12	15	-21.47%
Rice	5,911	1262.32	-0.9377	0.0115	1.7123	-0.0035	28	26	9.34%
Riley	33,784	828.67	-0.8534	0.0071	1.6470	0.0226	266	273	-2.72%
Rooks	3,458	1335.58	-1.1166	0.0210	1.8533	0.0166	20	13	55.94%
Rush	2,302	1219.82	-1.0884	0.0215	1.8197	0.0122	12	20	-39.92%
Russell	4,610	1303.30	-1.0219	0.0150	1.7838	0.0282	32	21	53.04%
Saline	37,729	1096.13	-0.8868	0.0073	1.6812	0.0024	236	262	-10.09%
Scott	3,317	776.41	-1.1976	0.0264	1.9111	0.0020	13	15	-15.47%
Sedgwick	350,456	1721.20	-0.9037	0.0109	1.6803	-0.0337	2,085	2273	-8.27%

Seward	12,163	721.67	-1.1730	0.0246	1.8932	0.0199	68	74	-8.05%
Shawnee	126,545	1012.08	-0.8454	0.0080	1.6342	-0.0039	803	785	2.23%
Sheridan	1,734	1246.75	-1.1952	0.0263	1.9095	0.0068	8	11	-24.95%
Sherman	3,187	1202.69	-1.2353	0.0279	1.9445	0.0248	19	10	89.38%
Smith	2,318	1418.25	-1.0642	0.0145	1.8321	0.0069	12	15	-19.65%
Stafford	2,626	1360.37	-0.9970	0.0178	1.7416	0.0057	14	12	15.32%
Stanton	1,184	785.07	-1.2118	0.0262	1.9272	-0.1253	2	0	N/A
Stevens	3,342	1033.85	-1.1904	0.0251	1.9092	0.0026	14	3	382.82%
Sumner	14,997	2048.75	-0.9177	0.0119	1.6893	-0.0133	79	80	-1.34%
Thomas	4,862	1488.99	-1.2174	0.0273	1.9284	0.0308	39	30	29.66%
Trego	2,096	1158.68	-1.1504	0.0242	1.8718	0.0323	13	14	-8.77%
Wabaunsee	4,729	904.41	-0.8505	0.0080	1.6395	0.0163	28	23	20.86%
Wallace	924	620.60	-1.2287	0.0275	1.9393	-0.0411	2	2	23.37%
Washington	3,909	1543.23	-0.8631	0.0062	1.6616	0.0288	28	22	28.79%
Wichita	1,233	828.91	-1.2132	0.0269	1.9255	0.0537	7	6	23.91%
Wilson	5,202	930.92	-0.8632	0.0100	1.6432	-0.0077	23	27	-15.04%
Woodson	1,908	749.09	-0.8587	0.0096	1.6405	-0.0040	9	10	-9.75%
Wyandotte	104,129	5.89	-0.8384	0.0082	1.6265	-0.0286	247	472	-47.59%
Total	1,971,987	114,797.26					10,883	11,212	